Ad Personalization and Transparency in Mobile Ecosystems: A Comparative Analysis of Google's and Apple's EU App Stores

David Breuer*
TU Darmstadt
Darmstadt, Germany
dbreuer@seemoo.tu-darmstadt.de

Lucas Becker*
TU Darmstadt
Darmstadt, Germany
lbecker@seemoo.tu-darmstadt.de

Matthias Hollick
TU Darmstadt
Darmstadt, Germany
mhollick@seemoo.tu-darmstadt.de

Abstract

Smartphones have become the primary interface to the Internet for many users, making app stores an essential part of the mobile ecosystem. Apple's App Store and Google's Play Store form a duopoly of the two largest app stores, both of which offer targeted advertisements in their store ecosystems. Consequently, their need for user data to improve the targeting of ads conflicts with users' desire for privacy. Users have to trust the statements given in privacy policies that are often scattered over multiple places and have no way of overseeing how their data is used for ad targeting. The European Union passed several regulations, most notably the DSA and DMA, addressing this transparency issue. The implementation of these laws, however, must be audited to ensure their effectiveness. Unfortunately, the transparency measures implemented in the context of advertising and the ad-targeting mechanisms in app stores have received little attention so far. In this work, we analyze the first-party ad tracking ecosystem on Apple's and Google's app stores. We measure the effects of different account parameters and interest patterns on the ads these accounts receive. Furthermore, we study the transparency measures implemented by the platforms. While we only detect rare occurrences of targeted advertising, we find Google's recommendations to be highly personalized. We notice multiple issues with the realization of transparency measures that affect their effectiveness and, in our opinion, contradict corresponding EU laws.

Keywords

First-party tracking, advertising transparency, targeted advertising, app store privacy, iOS, Android, DSA.

1 Introduction

App stores are the gateway to obtaining applications on mobile phones. As such, these app stores are an integral part of the mobile ecosystem. Users have to rely on them to install apps they need for their daily lives, such as banking apps or instant messengers. The global mobile app store ecosystem is divided into two major players, Google's Play Store (Android) and Apple's App Store (iOS) [91], despite the mobile app market being estimated at around 289 billion USD in 2024 [100]. Although there exist third-party stores for both operating systems, most apps are still installed via the platform's default app stores [90]. While Android allows to install apps without a store via sideloading, this workflow is inconvenient for daily use. Apple recently introduced a similar feature in iOS 17.5, reacting to the European Union's (EU's) introduction of the Digital Markets Act (DMA), but offering this only within the EU and limited to developers with at least one million app installs per year [21]. The resulting duopoly forces users to rely on the first-party app store of their chosen platform.

Consequently, Google and Apple possess substantial control over these ecosystems [33], and, by extension, over mobile users, which depend on their platform's app store to make full use of their devices. Both Google and Apple profit from their respective app stores by demanding service fees [16, 64] and selling in-store advertisements [15, 61]. For effective and hence most profitable advertising it is beneficial to target advertisements in order to reach the desired audience [97]. Such targeting requires information on the users and leads to a potential conflict of interest between the platform operator wishing to sell profitable advertisements and the user wanting to protect their privacy [98]. While both companies enforce rules on third-party app developers on how they are allowed to track a user's behavior to ship personalized advertisements [57, 67], these limitations do not apply to themselves. The resulting risks range from the manipulation of individual users by exploiting personal vulnerabilities [93] to large-scale, politically motivated campaigns [2, 76, 80]. Responding to these risks, lawmakers have passed laws to regulate and increase the transparency of online advertising systems. Most prominent in this context is the Digital Services Act (DSA) in the EU. However, to an outside observer, the mobile ecosystems of Google and Apple are highly complex black-boxes, despite these regulations aimed at increasing transparency. The DSA forces very large online platforms to publish repositories listing all distributed ads, but without trustworthy audits, the datasets published might not reflect reality, either intentional or due to implementation issues. For these reasons, we study the first-party tracking and advertising behavior in Google's and Apple's mobile app store ecosystems to increase transparency of how ad-targeting is applied and to gain insights into whether existing transparency measures are effective.

We therefore formulate the following main research questions:

- **RQ1:** How do Google and Apple personalize ads and recommendations in their mobile app stores? What are the differences?
- **RQ2:** Do Google's and Apple's app store advertising comply with the transparency measures of the EU's DSA?
- **RQ3:** Do users of Google's feature to exclude sensitive topics receive ads related to those sensitive categories?

*Both authors contributed equally to this research.

Proceedings on Privacy Enhancing Technologies 2026(1), 604–630

https://doi.org/10.56553/popets-2026-0031

Fundamentally, we address these questions by performing information flow experiments: We construct accounts with specific parameters or interests, so-called personas, and measure the ads displayed to them. We repeat these measurements to get a representative sample size and then evaluate them to deduce correlations between the persona we signaled towards the platform and the outputs (the ads) we received. We also use the privacy and personalization controls that Apple and Google offer to study their effects on the ads we observe. Further, we compare our results to the provided advertisement transparency repositories to analyze whether they are consistent with our experiments. Unfortunately, we find multiple implementation issues in the provided transparency measures, limiting the use of this comparison.

In summary, our paper claims the following main contributions:

- We present a methodology to examine black-box advertising in smartphone app stores, including an open-source automation framework. We focus our study on user targeting of app store ads and recommendations.
- We give insights into the advertisement ecosystems of Apple's App Store on iOS and Google's Play Store on Android. Our measurements amount to more than 4 million ads and recommendations from over 250 accounts.
- We describe inconsistencies in Google's app recommender system as part of the Play Store.
- We identify shortcomings of Apple's and Google's implementation of DSA advertisement transparency rules.
- We publish our source code and dataset to aid future research.

2 Related Work

In response to the opaqueness of tracking and targeting mechanisms employed by the advertising industry, researchers have been studying advertisement systems to uncover their behavior. A frequently used approach is to consider the respective advertisement ecosystem a black box, where only inputs (such as interactions with the system) and outputs (i.e., the resulting ads) can be observed. Researchers formulate insights into the tested black-box system by correlating inputs and corresponding outputs [73, 74, 85, 95].

Crowd-sourcing of ad targeting data. There are different approaches to constructing the dataset required for this process. One branch of research uses crowd-sourcing to record the interactions performed by real humans and the ads they are shown [45, 68, 71]. While crowd-sourcing enables the creation of large and diverse datasets that closely reflect the real-world usage of ad ecosystems, its effectiveness depends on the access to willing participants. Additionally, researchers have little control over the specific profiles (i.e., signaled interests and behavioral patterns) of users and cannot keep influencing factors fixed.

Persona construction. An alternative way to map inputs to outputs is to construct artificial personas that are used for interactions with the target system. Often used to detect ad-related tracking on the Internet, a typical pattern is to use a separate web browser profile for each persona. A persona is established by interacting with sites that are associated with a specific interest or characteristic and keeping the resulting client-side state. Later, a persona's state is used to measure outputs (i.e., ads) that can be linked to the previously signaled profile of the persona. Existing prior work targets primarily the web [1, 25, 26, 31, 34, 83]. Iqbal et al. [69] examine the Alexa Echo smart speaker ecosystem, although they still rely on the web for measuring ads. In contrast, we use personas to infer targeting mechanisms in the previously unstudied area of first-party tracking in mobile device app stores. Compared to the web advertising ecosystems studied in previous works, mobile first-party tracking happens within authenticated and closed ecosystems that introduce further challenges, which we discuss in Section 8.4.

Third-party tracking on mobile platforms. Advertisement-related tracking practices on mobile platforms have been studied before, often with a focus on advertising libraries shipped with apps [27, 72, 78, 88]. Works of a similar research direction have analyzed how apps can use permissions to leak sensitive data [30, 86] and how permission-related dialogues are perceived by users [82]. Our work is orthogonal to these works since we study tracking performed on the platform by first parties, which, due to their function, already are in possession of sufficient permissions to access sensitive data.

First-party tracking on mobile platforms. Prior work on first-party tracking on Android and iOS focused primarily on the data transmitted to Google and Apple [75], and other Android OEMs [79]. With respect to Apple's App Store, previous work found that user interactions in the store are logged and transmitted meticulously [84]. These studies cover only the network traffic sent to the first-parties, while we focus on the behavior of their black-box systems instead. Additionally, the association of search keywords and advertised apps has been investigated on iOS [94]. Here, the authors focused more on the relationship between competing apps and favored search keywords. We, in contrast, address questions related to user targeting and transparency.

Cross-device tracking. As we are limited to a fixed number of devices, cross-device tracking, in which advertisers link a user's different devices, is relevant when evaluating our data. Cross-device tracking on websites has been studied before [29, 89, 99]. The approach of Solomos et al. [89] is the closest to ours, since they also use artificial personas to detect targeted ads that result from cross-device tracking. In contrast to their work, we analyze mobile app store ecosystems and consider cross-device tracking as only one aspect of our analysis.

3 The mobile in-store advertisement ecosystem

In the following, we provide an overview of Apple's and Google's advertisement systems in their app stores by compiling publicly available information and introducing relevant EU regulations.

3.1 Advertiser View

Advertisers must create ad campaigns in order to get their ads shown in the app stores. Apple and Google offer web interfaces for this purpose. The option to set a maximum budget and bid is common to both interfaces. This is crucial, as the final decision on which fitting ads are shown to a user is based on an auction [17, 59]. In the following, we summarize the advertiser options presented on Apple's and Google's web interfaces.

Apple. Apple differentiates between two advertiser programs: *Apple Ads Basic* and *Apple Ads* (formerly Search Ads Basic and Search Ads Advanced). In the basic version, advertisers can only select a monthly budget, a maximum Cost-Per-Install (CPI) bid, and targeted countries. Apple claims to automatically choose the best targeting parameters per app and ads are only shown within the search results of the App Store [17]. Apple Ads allows developers to define their audience by a set of features: gender, age, device (e.g., iPhone or iPad), customer type, and location. Customer type allows targeting new users, users who have already downloaded the advertised app before, or users who installed other apps from the same developer. As a location, only wider areas can be selected, e.g., a larger city in Germany [23]. Advertisers can place their ads on the Today tab, Search tab, search results page, or product page [9]. If the search results are chosen, advertisers can additionally define a set of keywords that are matched with the search queries of App Store users, and define time spans over the day in which this ad campaign is active. Apple Ads is billed Cost-Per-Tap (CPT), where advertisers must pay for every click on an ad. The web UI offers no interest-based targeting options outside the search results placement [17].

Google. In contrast to Apple's UI, Google's advertiser web interface [61] offers detailed targeting options. Besides demographic features, including parental status and household income, Google allows targeting more fine-grained locations than Apple, down to rural towns in our experiments. Advertisers can choose from a large pool of specific interests or target groups that have experienced specific life events (e.g., graduation or marriage), and Google indicates whether these groups are willing to buy. Furthermore, advertisers can specify custom interest groups based on freely defined keywords, visited websites, or used apps. The personalization options exceed Apple's, thus confirming Google's core business model is advertising. To our knowledge, detailed options of Google's advertiser view are not publicly documented, but they publish data on targeting options as part of their real-time Bidding API [47].

3.2 Recommendations

Apple and Google operate recommender systems that suggest apps to users [12, 51]. It is opaque to users whether these systems are based on the same data as the advertisement systems, and if those are combined or separated. For recommendations, advertisers cannot specify any campaigns. The platform chooses recommended apps without any prior application or payment. Hence, these *recommendations* are not treated as advertisements, but have a similar effect on customers.

3.3 Privacy Policies

Apple and Google provide privacy policies in the EU as enforced by the General Data Protection Regulation (GDPR). In the following, we summarize their privacy practices regarding in-store advertisement and user tracking as stated in their privacy policies.

Apple. Apple's privacy policies are scattered over multiple locations. Apple provides a general privacy policy in combination with a summary [7, 22], individual privacy documents on every service [18], individual privacy labels for every app [14], and their Advertising Privacy Policy acting as a guideline for developers [4].

Apple distinguishes between ads based on *contextual information* and *profile information* [6]. The former is available at the moment the user opens the app store. This includes, e.g., the search term, the current content of the viewed page, the Operating System (OS) version, language settings, or the device's location. The latter is accumulated over time before the ad is served. In this process, the ad platform uses information such as store interaction, profile information such as name, age, or gender, and interaction with other apps controlled by the store operator.

For ads based on profiling, Apple groups users into segments of at least 5,000 people based on account information and interaction history with the App Store, News or Stocks app, and other Apple services like Music or Books. Apple Pay Transactions, Health app data, sexual orientation, religious beliefs, and political affiliations are excluded from their advertisement platform. Minors do not receive ads, and users can opt out of profiling-based ads in their App Store settings. Apple claims their "advertising platform does not track you" [6], which they define as not joining user data with data from other parties and not selling personal data to data brokers in the context of advertising.

Google. Similar to Apple, Google offers several websites with privacy statements. They provide a central privacy policy [51], multiple documents specifying their privacy practices for their individual services [3, 55], and information on their ad ecosystem for developers [48]. On their privacy web blog, they state their support in favor of privacy regulations [62]. They have also published a white paper on "Responsible Data Practices" [63].

According to their privacy policies, Google uses all data they collect from users and other parties to personalize recommendations and ads. They do not use data on health, race, religion, and sexual orientation for advertising, nor do they use data from Gmail, Google Drive, and Google Photos. Ads are not personalized if a user is underage. Every user can turn off the personalization of ads in their account settings. They define a list of sensitive categories consisting of alcohol, dating, gambling, pregnancy and parenting, and weight loss that users can exclude from their received ads [60]. They do not state what they regard as contextual and profile information.

3.4 EU Regulations

The most influential EU regulations affecting the smartphone app store ecosystem are the DSA [43], the DMA [42], and the GDPR [41]. In the following, we summarize the articles relevant to our work.

Digital Services Act. The DSA defines how online platforms within the EU must ensure consumer safety. It contains additional strict rules for Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs). These are platforms that have at least 45 million monthly users in the EU and are designated as VLOPs and VLOSEs by the European Commission. As Apple's App Store and Google Play are defined as VLOPs, all articles of the DSA must be applied [32]. Article 26 defines transparency rules for advertisements delivered through the respective platform. Information on who is responsible for an ad and why it is shown to the user must be available. Additionally, it prevents advertisements based on the profiling of certain personal features defined in Article

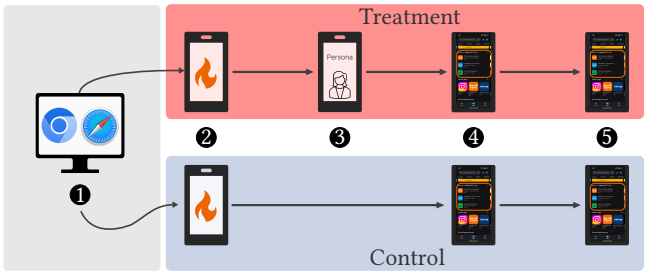


Figure 1: Measurement procedure of a single experiment pair: In Step ❶, we create accounts for each persona instance, and in Step ❷, we reset each device. During Step ❸, we install apps for our interest-based treatment personas. In Steps ❹ and ❺, we conduct our measurements before and after opting out of personalization.

9 GDPR, such as political opinion, religion, health data, and sexual orientation. Article 39 DSA extends this for VLOPs and forces platforms to provide the data through an online repository with API access, including information on how often an ad is delivered. Articles 27 and 38 define similar transparency rules for recommendations on the platform and ensure that consumers can opt-out of profiling-based recommendations. Article 28 prohibits the delivery of personalized ads to minors. Articles 34 and 35 define that VLOPs must assess their risks to society, report on that, and provide safeguards within their systems for risk prevention. Additionally, Article 14 defines transparency rules for a platform's terms and conditions, and Article 25 forbids manipulative online interfaces.

Digital Markets Act. The DMA focuses on fairness in competition between EU-defined *gatekeepers* and other businesses. For each gatekeeper, core platforms are identified, and Article 5(2) prohibits the combination of user data of a core platform and other services of a gatekeeper. As iOS, Apple's App Store, Google Play, and Google Android are defined as core platforms, Apple and Google are forbidden to combine user data collected by these services [33]. Article 5(2) can be relaxed through consumer consent.

General Data Protection Regulation. The GDPR defines boundaries for companies' data collection and strengthens consumers' rights against data misuse. It has been in force since 2018 and obliges companies to handle all user data transparently. Data minimization is emphasized, and data processing is limited to legal bases stated in Article 6. Additionally, it allows consumers to obtain a copy of their data and revise or delete it. One major improvement since its commencement is that companies now write more precise privacy policies and explain their data handling in greater detail [77].

4 Methodology

To address our research questions, we need insights into the functionality of the platform's infrastructure that go beyond the publicly available information. Fundamentally, this scenario can be considered as a black-box information flow problem [95], where we aim to measure the influence of information flows to Google's and Apple's app stores on the resulting in-store advertisements. We deliberately restrict our information flow problem to advertisements shown in the smartphone app stores, instead of also using advertisements from other services of the platform provider (e.g., YouTube in Google's case). The reason for this design decision is that we seek to compare Google's and Apple's ecosystems, meaning that we have to restrict ourselves to the subset of options available on both platforms. Since Apple at the time of our experiments only advertises in their app store [6], we focus on app store advertisements.

4.1 Overview

To measure the effect of information flows, we follow the approach outlined by Tschantz et al. [95] and always perform measurements twice: In the treatment group, we signal some property or behavior that we want to test (a so-called persona, see Section 5). In the control group, we repeat the same procedure but without signaling the behavior, as depicted in Figure 1. By this approach, we can compare the resulting output flows from the treatment to the control group. By simultaneously measuring the treatment and the control group (referred to as *measurement pair*), we limit the impact of temporal effects, such as different times of the day or shifts in the currently active advertisers.

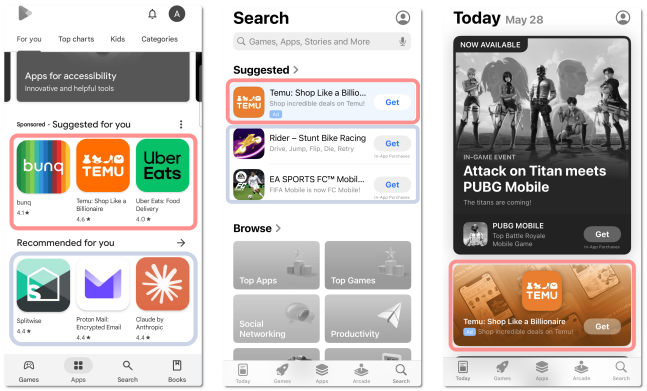
4.2 Experiment Procedure

In the following, we give an overview of the individual steps of each experiment, as shown in Figure 1. The technical details of these steps are explained in Section 4.4, and we explain our design choices in Section 4.5.

❶ *Account creation.* The first step of each experiment is the account creation phase. We create Apple and Google accounts in a separate environment that consists of a web browser and a smartphone (Samsung Galaxy S20 running Android 13). We use Chromium for Google accounts, and Safari and Firefox for Apple accounts. We partially automate this process using Selenium [87], but have to solve captchas manually to avoid bot detection mechanisms. The data used to populate these accounts is fixed, varied only according to the personas specified in Section 5. On both platforms, we enable all personalization settings where applicable, including Google's option to combine user data of all of their services [52].
We use eSIMs from a German service provider for the SMS-based verification step required during the account creation. Each account is tied to one eSIM. During account creation, we install these eSIMs sequentially on a separate Android smartphone dedicated solely to this task. This smartphone's internet access is limited to the API endpoints of our service provider that are necessary for eSIM activation. This way, we prevent any direct information leakage to either Apple or Google during the eSIM card installation. Such leakage could tie together different eSIMs and the singular device used to create accounts, tainting our experiments. The creation of Apple accounts requires an existing email address, for which we use email addresses reflecting the name of the created account hosted on our own email server.
After creating a new account, we wait at least three days before using it for our experiments. This time span is derived from Apple's documentation [8] and worked well for both Apple and Google accounts.

❷ *Device setup.* We fully reset each device at the beginning of an experiment. Then, we log in to the account under test.
❸ *Signaling.* The third phase of our experiments is the signaling phase. We install a set of apps with the respective account under test in case of our interest-based personas.
❹ *Measurement.* In our fourth experiment phase, we measure the ads our personas receive. We extract ads and recommendations on smartphones from Google's Play Store and Apple's App Store app. We use UI automation to open and scroll through the respective app store to achieve this. For each detected ad and recommendation, our setup writes the full name of the advertised app into our database. A background script on our server then fetches additional data regarding those apps directly from Apple's and Google's app store APIs. As we use the exact app store name of the examined apps, we can precisely match them through the store APIs. The retrieved metadata includes the app's store category, which is used later in our experiments.
❺ *Non-personalized measurements.* In phase five, we repeat measurements after turning off personalization to study the effects of opting out of personalized ads. This results in further subgroups within each measurement, distinguishing between the personalized and non-personalized subgroups.

Repetitions. We perform measurements of each persona multiple times with identical parameters and fresh accounts to get a more representative sample size. This leads to multiple accounts being associated with the same persona. The number of accounts per persona is listed in the Appendix in Table 2. While these accounts all entail the same persona characteristics, they do not reflect the same user. Each account is assigned a unique email address and SIM card. For example, we created our shopping persona five times for each platform, with individual accounts, email addresses, and SIM cards, so we can repeat the shopping measurement five times on each platform and increase our sample size.



(a) Google Play: Ads & recommendations **(b) Apple App Store: Search tab ad & recommendations** **(c) Apple App Store: Today tab ad**

Figure 2: Item containers we measure. *Ads* are marked in red, *recommendations* in blue.

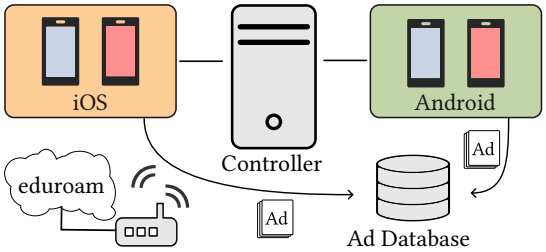


Figure 3: Overview of our experiment architecture.

4.3 Extracted Items

We extract the following ads and recommendations (referred to as items) from the respective app stores for our measurements.

Ads. Apple and Google show ads in various places in their respective app store apps on iOS or Android. Apple displays a single ad on the *Search* tab (see Figure 2b) and on the *Today* tab (see Figure 2c). Apple's App Store also displays an ad after performing a search and on an app's page, which we exclude due to their contextual nature, as these ads only appear after explicit user input. In Google's Play Store, ads on the main page are placed in scrollable containers marked as "Sponsored" (see Figure 2a), of which there can be multiple instances on the main page. Similar to Apple's App Store, the Play Store displays a single ad on the search page and an ad after performing a search. We ignore the former since we can get a higher volume of ads from the main page, and we exclude the latter because it is contextual.

Recommendations. In addition to ads, both platforms feature UI elements that recommend or "suggest" a selection of apps to the user. Unlike ads, these *recommendations* are not paid for by the app's developers and are not part of the bidding process. The app distribution of recommendations might differ substantially from the distribution of ads. This also means that the distribution of recommended apps is supposedly unrelated to advertising budgets and, therefore, less skewed towards apps with high advertising budgets. Their characteristics can, therefore, grant additional insights into the interest estimation of the platform operators.

4.4 Experiment Setup

To evaluate the results of our experiments, we need to control all user-facing inputs to the system under test [95]. In this section, we describe our test environment, discuss all experiment input parameters and explain how we handle them.

Figure 3 gives an overview of our experiment architecture. We use two Android smartphones (Samsung Galaxy S22+ and S23 running Android 14) and two iPhone 15 devices (running iOS 17.4) to run the related control and treatment experiment pairs in parallel. All smartphones are connected via USB to our experiment host (Mac Studio 2022 with macOS 14.5) running all our automation code. They are connected via WiFi to a router that forwards all traffic to and from a university-wide internet access network. Our setup blocks DNS requests to *gdmf.apple.com* to prevent the iPhones from automatically updating their operating system.

All information necessary to conduct the individual experiments, e.g., Apple and Google account login credentials, eSIM credentials, or persona details, as well as our experiment results, are stored in a PostgreSQL database that our experiment host accesses.

Physical devices. We use real devices to originate our information flows from. Better scaling alternatives would have been to use either emulated devices or to reverse engineer and re-implement relevant protocols (see for example [28]). We consider both of these alternatives infeasible. First, emulated devices are, without further tweaking, easily detected [70]. We cannot rule out that their use has a significant impact on the way that ads are distributed. Second, reverse engineering all relevant protocols to match real information flows with sufficient precision is labor-intensive and error-prone, given the large set of components involved in our experiments. Similar to using emulated devices, subtle differences to interactions performed by real devices might lead to dissimilar behavior. We shuffle devices between treatment and control groups to limit potential effects due to different device models.

Region and IP address. Our test devices are placed in a single building in Germany and are not moved during the experiments. We route our traffic through the university's network at the same location, thus achieving consistent location data via IP- and GPS-based localization. While this still allows grouping accounts by a common service provider, we found the alternative of using VPNs to mask our locations infeasible due to an increased occurrence of bot prevention checks [35]. Additionally, we share an IP address range with other legitimate device users of the university network.

Device automation. Our framework runs fully automated and does not require human interaction after initiation. We use the following tools to control the smartphones during our experiments.

On iOS, we use *cfgutil* to configure WiFi settings, execute a device reset, and perform the initial setup. This tool is part of *Apple Configurator* [13], which is designed to set up iPhones for companies. We have modified *cfgutil* to skip the manual Terms of Service agreement on the iPhones. We use *devmodectl* [20] to enable the developer mode on the iPhones after first start.

On Android, we use the AoAv2 protocol [56] to perform the initial setup and to enable developer mode by emulating a keyboard.

After the initial setup, we use XCode UI tests [24] to perform our experiments on the iPhones and Google UI Automator [66] to control the Android smartphones. We have written scripts to perform the following tasks on all devices through UI automation: *eSim insertion* and *ejection*, *Apple or Google account login* and *logout*, *app installation*, *applying privacy preferences*, and *ad measurement*.

4.5 Preliminary Experiments

We conducted preliminary experiments to tune our experiment's parameters to the system and devices under test.

Time between signal and measurement. For Android, we signaled an interest in finance by installing related apps, and repeatedly measure ads for a prolonged time of up to 20 days. In these experiments, we found that for Google's Play Store the initial time until a clear response to our signal was measurable amounted to roughly 10 days. We define a clear signal response as a visible change in the distribution of ad categories, e.g., visible between the two bars on the left of Figure 4a. Since we also observed that a strong signal was still present weeks after these 10 days passed, we decided for a "waiting" period of 14 days. For Apple's App Store, we found 7 days to be a sufficient waiting period until measuring.

On both platforms, we detected the effect of signaling the finance interest even if the associated account had been logged out of the device for the waiting period. This observation allows us to interleave signal and measurement steps of different experiments. In addition, we confirmed that the effect of opting out of personalized ads becomes visible almost immediately, which is why we only wait 10 minutes before starting with the non-personalized groups.

App interaction. We open each app briefly after installing to send a stronger signal since this action is transmitted to both Google and Apple. This is apparent in Google's Activity Centre [65], and in Apple's case, we observed this behavior in the network traffic of a jailbroken iPhone. As this can only be investigated on a jailbroken device and, while conducting this research, no publicly available jailbreak exists for iOS 17.4, we rely on data extracted from the network traffic of an iPhone 8 running iOS 16.7.8.

Amount of ads per time interval. We performed preliminary experiments to infer how many ads we can retrieve in a given period before running into rate limits of the respective store. Google's Play Store presents users with more ads than Apple's App Store. This means that by scrolling through the front page without resetting the app's state, we observe 126 ads on one page of the Play Store, while we can retrieve only two ads from the App Store without restarting the app: One ad on the Today tab and one ad on the Search tab. We re-open the respective app store to retrieve additional ads. However, after measuring ads for a prolonged time, the number of ads we can get drops drastically, falling to around 12 ads per iteration on the Play Store and no ads at all on the App Store.

For this reason, we throttled our measurement process. We extract (in total) at least 252 ads per measurement run from the Play Store and 120 ads from Apple's App Store. After extracting the targeted number of ads, our script waits until a full hour has passed since the beginning of the measurement run to account for rate limits before starting the next measurement iteration. While we could have throttled both platforms to the same amount of extracted ads, we decided to extract as many ads as possible per platform to strengthen the validity of each platform's experiments. Due to ad-availability patterns in our preliminary experiments, we run the measurement process on Android three times and two times on iOS for each account, resulting in extraction times of three hours for Android devices and two hours for iPhones. We measure only two times on iOS since the number of ads fluctuates heavily in the third measurement run. These fluctuations prolong the time needed to get 120 additional ads by a non-deterministic amount, thus making it infeasible to schedule experiments properly.

Cross-account effects. Since we only have a limited number of devices for experiments, reusing them might lead to two issues.

First, Apple and Google could transfer the interests of one account to a subsequently logged-in account, tainting the measurement. To study this behavior, we perform additional tests, where we log in to two accounts (control/treatment) on two devices in

parallel, where we measure a clear difference between control and treatment after signaling. Then, we log in to two accounts associated with neutral personas and compare them. We did not detect any differences between those two accounts, which resemble the difference between the former two accounts, and hence, we conclude that no measurable cross-account effects exist. Nevertheless, we reset all devices before logging in to Apple or Google accounts.

Second, Apple and Google could use the device type to assign a specific interest category, leading to different results depending on the device on which an account is logged in. We analyze the results from our baseline experiment to examine such a potential behavior. We group our extracted data by device and search for patterns where repetitions of certain ads or recommendations on the same device are closer than repetitions on the other. Here, we again found no such pattern for either platform.

5 Persona Design

Personas reflect the input parameters to our black-box information flow experiment. We assign them to all our accounts and limit each persona to one modified test input. All other experiment parameters remain unchanged to ensure the outcome of our test originates from our test input. We divide our personas into three types based on their test input: *Neutral*, *Account Parameters*, or *Interest-based*. Their exact parameters are specified in Table 3 in the Appendix.

5.1 Neutral Persona

We use neutral personas for the control groups and for a baseline experiment comparing two control groups to each other. We assign these personas common values such as names to make them a fitting representation of a typical user within Germany. The account-parameter-based and interest-based personas use the values of the neutral persona for all parameters except the one being tested.

When creating Apple or Google accounts for our personas, we must specify static account parameters, including name, age, email, and gender. However, there are slight differences between Apple and Google. For an Apple account, we need to provide a valid email address, while Google offers to create a Gmail address for each Google account. The email prefix is always constructed out of the first and last name of the account and a six-digit number. We use *Alex Müller* as the neutral account name. It consists of the most common names in Germany [92] and is gender neutral. There is no option to provide a gender when creating an Apple account. For Google Accounts, a gender must be selected, so we chose the option *prefer not to say* as the most neutral option. For the mandatory birth field, we use the first of January 1990, which results in an age of 34 years at the time of our experiments. Our neutral persona does not install any app from the app store.

5.2 Account Parameters Persona

Both companies use account information for personalization according to their privacy policies [6, 51]. For this reason, our second type of personas examines the influence of different ages, genders, and names on the examined advertisement platforms.

Age. We split our age personas into three clusters: Young adults (25 years), middle-aged adults (50 years), and seniors (85 years). We choose these clusters to cover a broad range of different generations. Apple and Google do not serve ads to minors [6, 54]. Therefore, we do not create underage personas. During our preliminary experiments on Google's Play Store some accounts were flagged as minors and did not receive ads. Hence, we can confirm this behavior.

Gender. Apple states to derive a gender from the account name [6], and we assume a similar behavior by Google. To inspect this behavior, we use the account name parameter to signal a gender to the platform while still using Google's gender option *prefer not to say*. We design our gender personas by using the most common first names in Germany for people born in the year 2000 for a given gender: *Maria, Sophia, Julia, Laura, Anna, Alexander, Maximilian, Lukas, Philipp*, and *Daniel* [92]. The last name is always *Müller*.

Name. We want to test the influence of common names of other cultures on the received ads and recommendations. For this, we use the most common names in Turkey and China as representatives for large non-European cultural groups in Germany [44]: *Mehmet Yilmaz* and *Zhang Wei* [81, 96].

5.3 Interest-Based Persona

For our third type, personas are defined by their behavior instead of their static account parameters. We restrict our interaction with the advertisement platforms to the app stores for comparability reasons (see Section 4). Therefore, the behavioral input to our experiments is interactions within the app stores, i.e., app store search queries, app installations, and app launches via the app store UI. Apps shipped with the OSs are not installed via an app store, and we assume these apps have no impact on our experiments. Even if these apps were reported to the store operator, they would be present in all experiments and, therefore, be part of our baseline.

We assign a set of 30 apps to each persona. These apps are installed and opened during the signaling step. We create a set of apps for all personas using the top free app charts of the respective store categories, where possible. For personas not directly mapped to an app store category, we curate a list of free apps that reflect the persona's interest. For example, we search for apps that track food consumption to reflect the interest of our weight loss persona and add the top search results to the respective app list.

First, we curate two personas that signal generic interests: *Finance* and *Shopping*. These personas are intended to reflect typical user interests combined with a strong monetary incentive on the advertiser's side, and should, according to our intuition, produce a strong effect. Our other interest-based personas are chosen to reflect topics that are linked to potential risks as outlined by the DSA. For these, we pick *Alcohol Sobriety, Parenting, Gambling, Weight Loss*, and *Dating*. These are the topics that Google deems sensitive, offering the option to opt out of receiving associated ads (see Section 3.3). We exclude the *Dating* interest due to the need for age verification when installing adult-targeted apps. We test this option by marking the corresponding topic in these personas' Google account settings as sensitive.[1] Finally, we design a *Mental Health* persona according to Article 34 DSA.

[1]We exclude the sobriety persona from this process, which was not marked as sensitive due to a malfunction of our tooling.

Table 1: Permutation test results with chi-squared as the test statistic and 9,999 performed permutations (R = 9,999). We use the category distribution of measurement pairs as input (C = Control, T = Treatment, P = Personalized, NP = Non-Personalized). Significant pairs are written first, with the number of insignificant pairs in parentheses ($\alpha = 0.05$). Sum of Play Store recommendations differs from the sum of ads due to missing recommendation data. Stable results are marked in bold.

Platform	Baseline				Account Parameters				Interest Groups			
	NP/NP	P/NP-C	P/NP-T	P/P	NP/NP	P/NP-C	P/NP-T	P/P	NP/NP	P/NP-C	P/NP-T	P/P
Play Store (Ads)	**5** (0)	**5** (0)	**5** (0)	**5** (0)	19 (3)	**22** (0)	21 (1)	**22** (0)	33 (2)	34 (1)	33 (2)	**35** (0)
Play Store (Rec.)	3 (2)	**0** (5)	**0** (5)	1 (4)	10 (9)	6 (13)	6 (13)	7 (12)	**34** (0)	1 (33)	4 (30)	**34** (0)
App Store (All)	**0** (5)	**0** (5)	**0** (5)	**0** (5)	**0** (23)	1 (22)	1 (22)	**0** (23)	**0** (34)	**0** (34)	**0** (34)	**0** (34)
App Store (Search tab)	**0** (5)	**0** (5)	**0** (5)	**0** (5)	**0** (23)	1 (22)	1 (22)	**0** (23)	**0** (34)	1 (33)	**0** (34)	**0** (34)
App Store (Today tab)	**0** (5)	**0** (5)	**0** (5)	**0** (5)	**0** (23)	**0** (23)	**0** (23)	**0** (23)	**0** (34)	**0** (34)	**0** (34)	**0** (34)

6 Evaluation

In our evaluation, we analyze the extracted *ads* and *recommendations* (both referred to as *items*) regarding their *category distribution* and *frequency*. We interpret our data using statistical means of *permutation testing* and the *Jaccard similarity index*. Finally, we manually inspect our data to uncover further peculiarities.

We explain our key evaluation methodologies in the following paragraphs and provide detailed results of the two examined platforms in Sections 6.1 and 6.2.

Category distribution. We group the extracted items of each experiment by their respective store category and calculate the *category distribution* for each experiment. Figure 4 shows exemplary results of one *Shopping* persona experiment. Apple and Google state categories for each app within their stores, e.g., *Finance* or *Shopping*, and we fetched those during our experiments (see Section 4.2). To simplify the interpretation, we group all *Gaming*-related categories into one group.

Frequency. We calculate the *frequency* of extracted items by counting all occurrences of *ads* or *recommendations* of each experiment. We identify all items by their name (see Section 4.2). Figure 5 shows the frequency of extracted ads over multiple experiments.

Permutation testing. We follow the methodology of [95] and use permutation testing to detect significant differences between our experiment groups. Our test statistic is chi-square. In contrast to Pearson's chi-square test, permutation tests do not assume the population distribution under test [46]. We execute 9,999 permutations for each test and use a significance level of 0.05. We use the category distribution of ads or recommendations for these tests. In Table 1, we list the number of significant experiment pairs.

Jaccard similarity index. We use the Jaccard index to inspect our frequency data and detect trends in the similarity of experiment pairs that do not show significant differences according to our permutation tests. A low Jaccard index indicates a low similarity.

Dataset. In total, we extracted 257,820 ads and 378,057 recommendations as part of our main experiments. We created 249 Apple or Google Accounts using 213 eSIMs. We conducted our experiments from December 2024 to January 2025. In conjunction with our preliminary experiments, we extracted 1,130,243 ads and 3,514,632 recommendations with 362 Accounts using 281 eSIMs.

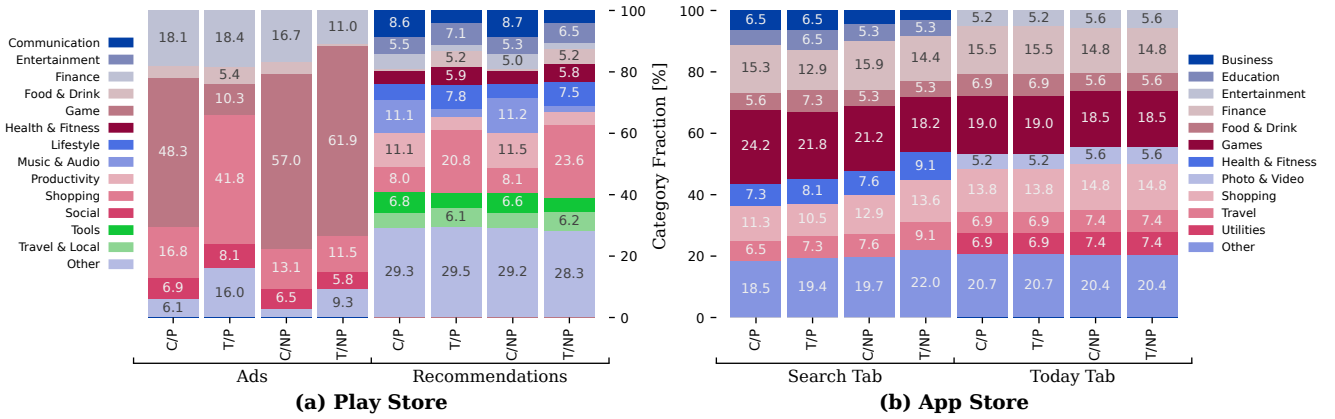
6.1 Google's Play Store

In this subsection, we analyze the data measured on Android smartphones. On Android, we observe considerable differences between the personalization of *ads* and *recommendations*. For this reason, Android-related tables and figures are arranged by this aspect.

Inconclusive permutation test results for ads. Our permutation test results for the *category distribution* of *ads*, depicted in the first row of Table 1, consistently indicate that samples are drawn from different distributions. Here, the three persona types exhibit similar characteristics, showing no differences in personalization patterns of the baseline, interest groups, and account-parameter-based personas. However, our baseline persona uses the same neutral persona type for control and treatment. Since the baseline always produces significant permutation tests, we conclude that the test results are unsuited to detect personalization effects in this setting.

Personalization settings do not affect recommendations. The results for *recommendations* differ from those for *ads*. For the baseline persona, the majority of measurement pairs produce insignificant tests, so we cannot prove that they are drawn from different distributions. This result matches our intuitive expectation for the baseline persona. The tests for the account-parameter-based persona result in a mix of significant and insignificant cases. The test results for the interest-based persona are most pronounced. Here, the differences between control and treatment are always significant. But, when comparing personalized to non-personalized pairs within treatment or control groups, the test results are mostly insignificant (see Table 1). These results imply that disabling personalization does not affect the *recommendations*. This effect is also visible in the category distribution of recommendations (see Figure 4a).

Persona-induced personalization effects. We observe multiple instances of personalization effects indicated by changes in the distribution of categories. The most pronounced case for *ads* is in the *Shopping* persona, where three out of five treatment measurements show strongly increased frequency of *ads* from the shopping category compared to the control group (see excerpt in Figure 4a and full Figure 8h in the Appendix). The Chinese name persona of the account-parameter type shows an increased number of games in three out of five treatment measurements, while there is no such pattern for the Turkish name persona. All other personas do not show apparent differences between control and treatment in the category distribution of *ads*.



(a) Play Store **(b) App Store**

Figure 4: Category distribution of one measurement pair (C = Control, T = Treatment) of the shopping persona before and after deactivating personalized ads (P = Personalized, NP = Non-Personalized). Full data in the Appendix, Figures 8h and 9h.

For *recommendations*, the effects of our signaling procedure have pronounced effects on the category distribution for most interest-based personas. These effects are visible in Figures 8h to 8l in the Appendix, where the frequency of a specific category starkly increases compared to the control group. This increase is consistent within all measurements and persists after opting out of personalization. The permutation tests support this observation since they are always significant between control and treatment for interest-based personas, and they also indicate that these effects stay after disabling personalization. Another observation for *recommendations* is that the interest-based personas *Gambling* and *Sobriety* (see Appendix, Figures 8m and 8n) only show subtle changes between control and treatment. This difference from the other interest-based personas could be due to the fact that these personas do not have a clear mapping to a single app category.

We further quantify these results by comparing the Jaccard similarity of the extracted items (see Appendix, Figure 6). For both *ads* (left-hand column) and *recommendations* (right-hand column) the Jaccard similarity indicates the same characteristics observed by the *category* distribution. For *ads*, this means no clear difference between different groups, while *recommendations* show a pronounced difference between control and treatment for interest-based personas (see Appendix, Figure 6f), but not for account-parameter-based personas or the baseline.

Ad campaigns. Figure 5a shows how often a specific *ad* occurs in each measurement. It highlights that some ads occur in almost every measurement, while the majority only appears in some of them. Sorted by their total number of appearances, their frequency gradually drops down to single occurrences over our dataset. This characteristic implies that a small set of developers allocates a large enough advertising budget to steadily be included in our measurements, while most developers pay for a lesser number of impressions. For *recommendations*, the distribution is similarly, but with spikes of apps that only appear in specific personas (see Figure 5b).

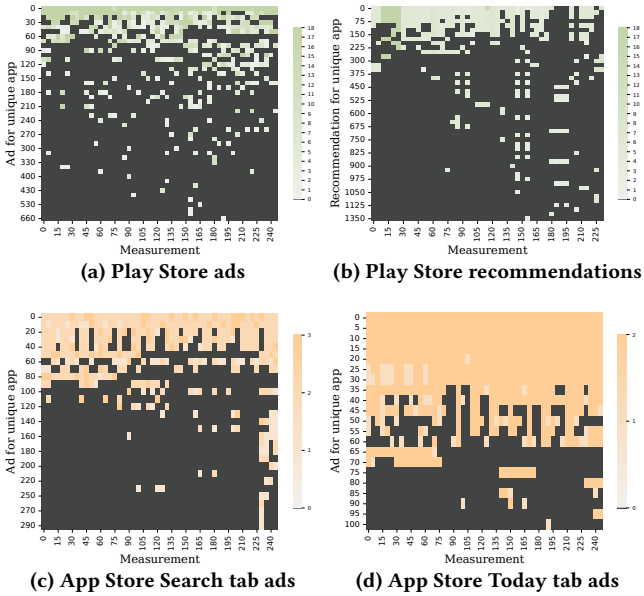
Unique items. When looking at *ads* that appear uniquely in a single experiment group, we notice the following: Our shopping persona got a selection of nine different shopping or fashion ads that exclusively appeared in the treatment group. These appeared in three of the five treatment extractions. This is consistent with our observation that the shopping persona is the only instance with visible personalization effects. More interestingly, the parenting persona got four unique ads that have kids as their primary audience: 'Sago Mini World: Kids Games', 'Montessori Preschool, kids 3-7', 'Sago Mini School (Kids 2-5)', and one toy-related ad: 'LEGO® Builder'. These are all concentrated in a single measurement, which leads to the assumption that this specific account has been assigned a "parenting" interest[2]. The Gambling treatment group has one instance with eleven unique game ads, compared to 0.3 unique game ads per instance on average. Hence, we hypothesize that our gambling persona has been categorized as "mobile gaming" interest.

There is a higher amount of *recommendations* that uniquely appear in single experiment groups compared to *ads*. Some frequent examples include: Anxiety and AI therapist apps that are unique to the mental health persona, baby monitor and apps targeted toward kids for the parenting persona, fasting and workout apps for the weight-loss persona, and various habit trackers for our sobriety persona. This degree of personalization seems to imply that Google is able to correctly identify the interests of these personas, even if their classification does not always match our intention for a persona (e.g., mobile gaming instead of gambling). It seems that Google does not directly apply this information for the targeting of *ads*, since we did not measure similar effects there.

6.2 Apple's App Store

Our results from Apple's App Store show less clear personalization patterns than those from Google's Play Store. We observe a steady distribution of ads along the App Store categories throughout all experiments. While the results fluctuate, there are only little signs of

[2]Another potential explanation is that marking the "pregnancy and parenting" topic as sensitive was unsuccessful for this instance, either due to an issue with our tooling or because of an error on Google's side. We noticed multiple cases where the MyAd-Center UI was slightly broken and options were missing. Unfortunately, opting out of personalization resets sensitive topics, so we could not verify this setting in retrospect.



(a) Play Store ads (b) Play Store recommendations

(c) App Store Search tab ads (d) App Store Today tab ads

Figure 5: Heatmaps that show how often items occur per measurement (frequency). Rows represent an ad or recommendation for a specific app. Columns are individual measurements sorted by time. Rows and columns are evenly selected over the available data. Rows that are all-zero because of this sampling method are omitted.

our personas being targeted. This is supported by our permutation tests that indicate a significant difference between merely four personalized to non-personalized pairs within the ads extracted from the Search tab (see Table 1). Looking at Figure 4b, this behavior can be observed in our category distribution by the example of our shopping persona. A slight difference is visible between the control and treatment groups of the Search tab, while a higher difference is visible between the personalized and non-personalized groups.

We assume that this behavior is caused by two factors: The limited number of ads we can measure in a feasible amount of time and a low degree of personalization.

No personalization of recommendations. We detect no variation between the individual measurements of recommendations. The recommendations consist mainly of mobile games across all experiments. Because of that, we exclude them from our evaluation of Apple's App Store and focus on the differences between the ad placement options Apple offers on their iOS App Store, which are the Today and the Search tab.

Differences between ad placement options. First, we examine the differences between iOS' two ad placement options, the *Today tab* and the *Search tab*. Our permutation tests do not show any significant differences in the ad distributions extracted from the Today tab (see Table 1). The mean Jaccard similarity between the personalized treatment and control groups is consistently higher on the Today tab than on the Search tab with all our persona types (see Appendix, Figure 7). The equivalent non-personalized treatment and control groups show similar behavior. We constantly measure comparatively low Jaccard indices between the personalized and non-personalized groups.

We conclude that a basic level of personalization exists that is higher on the Search tab than on the Today tab. We consistently measure differences between groups with and without active personalization. The difference between our control and treatment groups is very low compared to the aforementioned personalized and non-personalized groups, and there are multiple explanations for this behavior. On the one hand, the measurable sample size might be too small, and on the other hand, the personalization on Apple's App Store might be very subtle.

Figure 4b shows these results using the example of the category distribution of one of our *Shopping* persona experiments. On the Today tab, the distribution is the same between measurement pairs. On the Search tab, the distribution differs slightly between measurements, but no clear pattern is evident. The most apparent effect is visible between the Search and Today tab ad distribution, which indicates the ads of the two placement options stemming from different populations. This effect is evident consistently throughout our experiments (see Appendix, Figure 9).

Less ad diversity on Today tab. We observe frequent drops in the measurement rate on the Today tab during our experiments. After an unspecific amount of time, we do not receive any ads on the Today tab. We do not get an ad more than two times within one measurement run on the Today tab and not more than three times on the Search tab within two hours. In conjunction with that, we measure fewer distinct ads on the Today tab than the Search tab.

We assume a small set of large companies compete for Today tab ad placements. Smaller companies may be unable to compete with these companies regarding advertisement budgets, and their ads are less often shown here. There seems to be a rate limit by Apple that every ad is only shown twice within two hours. We assume that this policy is in place to ensure the diversification of ads on the Today tab. Nevertheless, it is ineffective, as only large ad campaigns are shown on the Today tab.

A subset of these large *ad campaigns* is visible in Figures 5c and 5d. In total, we observed a group of 45 ads throughout our experiment period in every measurement. The majority of these reoccurring ads are shown on the Today tab.

High personalization impact of account parameters. We experience more distinct results from experiments involving account parameters such as gender and age (see Appendix, Figures 7c and 7d). Our interest-based persona experiments do not yield conclusive results but fluctuate strongly across calculated category distributions. This behavior correlates to Apple's advertiser web UI, which presents no options for interest-based targeting on the Today and the Search tab. Still, we measure a light amount of interest-based personalization from Apple's opaque targeting mechanics.

In our gender group, we observe the app *BeTidy: Cleaning Schedule* in three out of five female accounts. Apps like *Business Card Scanner by Covve*, *OANDA: Forex & Stocks Trading*, or *ONLOGIST | Drive & earn money* are only observed through our male accounts. We do not observe effects based on culturally significant names.

Baseline experiment. Our iOS baseline experiment shows a low mean Jaccard similarity between the personalized and non-personalized pairs. Additionally, we find a high similarity between control and treatment groups (see Appendix, Figure 7a). We observe a higher frequency of gaming ads and a lower frequency of finance apps in the non-personalized group. The clear distinction between accounts with active and inactive personalization indicates general ad personalization on Apple's App Store and a measurable effect of the corresponding opt-out functionality, consistent with our aforementioned results.

7 Compliance and Transparency Issues

In this section, we discuss compliance issues in relation to the DSA as well as other transparency problems related to **RQ2**.

7.1 Ad Transparency Datasets

Per Article 39 of the DSA, Google and Apple must provide ad transparency datasets for advertising within the EU. We initially intended to refer to these datasets as ground truth to evaluate our experiments but found them affected by several issues.

Google's Ad Transparency Center. Google's Ads Transparency Center can be accessed using either the web interface [53] or Google's BigQuery cloud infrastructure [58]. Access using the cloud API allows for arbitrary complex queries, which exceeds the simple filters that are possible using the web interface.

The data shown on the web interface seems incomplete compared to the records in the BigQuery dataset. We noticed multiple instances where the web interface does not show any text ads for an advertiser in the Play Store. The dataset, however, lists multiple matching entries. The *creative_page_url* field, which usually links to the respective page on the web interface, resolves to a page-not-found error in these entries.

The dataset schema defines a field called *surface_serving_stats* to distinguish "the platform where the ad served [sic]. Possible values are: SEARCH, MAPS, PLAY, SHOPPING, YOUTUBE" [50]. These values correspond to the options in the web interface.

When comparing the observable ads in Google Play with the ads in the Google Ad Transparency Center, it becomes evident that the advertisements displayed in Google's Play Store are not represented in the Ad Transparency dataset. We found multiple instances where apps that we encountered in the Play Store did not appear in a matching format in the Ad Transparency Center. When looking up ads attributed to the corresponding apps and the related advertisers in the Ads Transparency Center, only video ads are shown, while filtering for either text or image ads yielded no results. However, the ads displayed in the app store are clearly not video ads since they only display a static image accompanied by text (see Figure 2a). Manual inspection of these video ads reveals that they do not correspond to the ads seen in Google's Play Store. In addition, we obtain the same results when directly querying the downloaded dataset. We hence conclude that the ads shown in Google's Play Store are entirely missing from the Ad Transparency dataset, which contradicts Article 39 DSA, and we suspect that instead, these ads are part of Google's AdMob Service and are shown in third-party apps [49].

Apple's Ad Repository. Similar to Google, Apple provides a web interface [10] and an API [11] for its Ad Repository including documentation [5]. We did not detect missing entries by comparing randomly selected samples of our dataset to Apple's Ad Repository. However, the provided details of an ad do not contain any information on the number of service recipients. This information is explicitly required as part of the ad repositories by Article 39(2), point (g) DSA. However, Apple only lists the aggregated numbers of total service recipients per country as part of their transparency report [19]. While Article 39(2), point (g) is not precisely clear, if the recipients must be listed per ad, we argue that Article 39(1) refers to"an advertisement" and extends Article 26, which refers to individual ads. Additionally, it would be of more value to the public and our research community to gain knowledge about individual ads and in line with Recital 95 DSA, which states that ad repositories should "facilitate supervision and research into emerging risks brought about by the distribution of advertising online."

7.2 Store UI Issues

In this subsection, we discuss two cases where missing information from the app stores leads to reduced advertising transparency.

Unmarked contextual ads in the Play Store. When performing a search in the Play Store, a single ad is shown at the top of the result list. Semantically, this ad matches the executed query. However, in all observed cases, these ads were never marked as contextual in the "Why you're seeing this ad" info but only as a product of the time and day and the general location. We argue that this conflicts with Article 26(1), point (d), which requires "meaningful information" to be presented "directly and easily accessible from the advertisement." Recital 68 clarifies that contextual ads should be marked as such.

Missing advertiser name in Apple's App Store. In Apple's App Store, the "About this Ad" section states targeting information associated with the ad. However, this page does not contain information on the advertisers themselves, only a link to the app's store page. Since the store page does not clarify the natural or legal person on whose behalf this ad is presented or from whom it is paid, we argue that this violates Article 26(1), point (b) DSA. Apple claims that the ad is always displayed on behalf of the app developer. Still, this fact is stated implicitly and not in a "clear, concise and unambiguous" fashion to "each individual recipient" as Article 26 DSA requires.

8 Discussion

The results of our analysis show interesting differences between Apple's and Google's app store ecosystems. Furthermore, we detected multiple transparency issues of these systems in regard to the DSA. In the following, we answer **RQ1** to **RQ3** by combining our experiment results with our analysis of the app store ecosystems.

8.1 RQ1 - Mobile App Store Personalization

Apple and Google differ unambiguously in the amount of ads within their app stores and the intensity of personalization. In this section, we compare Apple's and Google's advertisement approaches.

Privacy policies. Both companies do not present their policies in a single document but scatter privacy-relevant information over multiple documents. This appears to be confusing from a user's

perspective and leads to misleading privacy assumptions, and we recommend that companies avoid this.

A difference lies within the data sources used for personalization. Google operates multiple platforms that are used as input to its advertisement system, such as its search engine, YouTube, Ads, Android, and Maps [51]. While a general combination of this data is not allowed under Article 5(2) DMA (see Section 3.4), Google frequently asks its users to agree to link their services. While we expect most users to consent to this at some point, the general process of how companies obtain user consent is not within the scope of this study and should be studied further.

While Apple operates multiple services themselves too, they constrain the input to their advertisement system to a subset of their services (see Section 3.3). In addition, they do not operate services similar to Google Search, YouTube, and Ads.

Advertiser view. Our analysis of the advertiser user interface confirms the differences in their privacy policies. The level of detail for targeting specific user groups is exceptionally high through Google's interface. In contrast, Apple only allows targeting broad demographic features. Nevertheless, it remains opaque how Apple uses the data stated in their privacy policy for personalized advertisements because the advertiser options do not reflect these options. Interestingly, Apple states in their advertiser documentation that 78% of users did opt out of personalized advertising in 2024, and ads that target demographic features are not shown to users who did opt out [23]. This might encourage more advertisers to favor contextual ad campaigns over personalized ones.

Intensity of personalization. Our experiments show differences in the personalization intensity of the compared systems. We observe the strongest personalization effects in Google's recommender system. With the same accounts, we observe fewer ads that stem from personalized content. We assume this is caused by the distribution of advertisement budgets. While recommendations are not paid for by app developers, Google only displays ads from paying advertisers. This reduces the pool of eligible ads and also skews the ad distribution towards apps and categories that have a profitable business model. Simultaneously, smaller advertisers with limited budgets compete with large companies, which leads to a concentration of large apps being advertised prominently.

This effect is even more explicit in Apple's App Store. Regardless of the persona type, we consistently observe the same large apps advertised across all experiments. Due to the fewer personalization and ad placement options, we assume the competition between advertisers is higher than in Google's Play Store. Ads from smaller advertisers are only shown on the Search tab after a prolonged measurement time, which are infeasible to extract.

We observe different personalization effects depending on the persona type. On Apple's App Store, we measure a higher impact of account parameters on personalized ads than our signaled interest. On Google's Play Store, this effect is reversed, which corresponds to our insights into the respective platforms' privacy policies and advertiser views. Regarding the influence of culturally significant names on targeting, we only observe more gaming ads for our Chinese persona in Google's Play Store.

Temporal effects. We observe strong temporal effects across all our experiments. The distribution of ads is dependent on the respective time-bound ad campaigns. Successive experiments were more likely to receive the same ad than other experiments. However, the distribution of the ad categories stays approximately the same.

8.2 RQ2 - DSA Compliance

During our research, we detect DSA infringements regarding advertisement transparency. We have identified a discrepancy between certain DSA articles and the implementations of Apple and Google (see Section 7). Both companies have been fined by the European Commission several times in the past for antitrust reasons or abusive app store rules [36–40].

While we strongly appreciate the effort of the DSA to enforce advertisement transparency, we notice a current research gap in the monitoring of closed platforms concerning compliance with EU regulations. In conjunction with independent researchers, EU agencies should oversee the platform's advertisement and recommender systems and constantly improve regulations. With this work, we hope to provide methodology and tooling to inspire fellow researchers to inspect locked-down systems.

8.3 RQ3 - Sensitive Topics

During our experiments, we found no indications that Google violates its rules regarding ads shown to sensitive personas. Users can mark a set of topics as sensitive and do not receive any related ads.

Nevertheless, our analysis proves this does not apply to recommendations that are not marked as advertisements. Our sensitive personas received a significant number of recommendations for apps that are subject to a sensitive topic. For example, our weight loss persona received recommendations for dieting and fitness apps while receiving no ads for such apps. While this does not directly infringe on their privacy policy, we are convinced that a user can reasonably expect not to get any recommendations or ads for topics that are explicitly marked as sensitive. This behavior can lead to "negative consequences to the person's physical and mental well-being" [43] applicable to Article 34 of the DSA and, thus, should be treated as a potential systemic risk. Although Apple does not allow its customers to mark specific topics as sensitive, we conducted the same experiments on iOS. However, we could not detect any patterns that indicate targeting of sensitive topics.

8.4 Hurdles Analyzing First-Party Ecosystems

We encountered several systemic obstacles while conducting this research, originating from the opaqueness of the platforms.

First, we had to acquire hundreds of phone numbers, including SIM cards, to create a sufficient number of accounts. This necessity increases the effort required to research the store ecosystems. Moreover, some potentially interesting app categories, such as dating apps, require age verification to install them. We could not provide any of the available verification measures at the required scale.

Secondly, we needed to design and implement a system to automate our experiments on two distinct platforms. The respective app stores are deeply integrated into the OSs and transmit several device identifiers in each HTTP request [75]. Because of this, we refrain from obtaining the ads purely through API calls or from an

emulated device, as we could not ensure the validity of our data. The notable disadvantages of this approach are that our experiments are expensive to scale due to the requirement for physical devices and that we have to reuse mobile phones for multiple personas. Since device identifiers are sent to Apple and Google, they are able to track this reuse of devices. Our preliminary experiments tested for such cross-account effects and failed to measure any. Therefore, we assume that they refrain from cross-account tracking for ad targeting, as it would endanger users' privacy when re-selling used devices. Nevertheless, we shuffle devices between measurements to account for potential device effects on our measurements. In addition, frequent but subtle UI changes in the app stores lead to issues that have to be resolved manually.

Thirdly, we experienced severe rate limits and banning of our test accounts, limiting the number of ads we could measure on each platform. Google started to require captchas to log in to our accounts, increasing the manual effort for our study. Although we successfully requested the re-activation of banned accounts, we cannot rule out that they have been tainted by this process. Overall, these obstacles resulted in smaller datasets, especially for iOS, and less generalizable statements on the tracking ecosystems of Apple and Google. Furthermore, fraud detection algorithms could influence our measurements to deviate from a normal user's view.

9 Limitations and Future Work

We use one phone number for one Apple and one Google account to maximize the number of accounts we can create. Hence, we have to assume that Apple and Google do not collaborate to track phone numbers beyond their own ecosystems. All the email addresses used to create Apple accounts share the same domain. It is common practice for people to share the same email provider, so we do not assume that Apple links our accounts by their email domain.

We also perform all our measurements from the same geographical location, sharing the IP range of a university. Although we think that this is necessary to prevent location-based effects in our setting, follow-up works could study the effects the location has on ad targeting in app stores. For example, platform operators could assume that we are students or university employees when creating the interest profiles of our accounts.

If Apple and Google address the shortcomings in their ad repositories (see Section 7), it could be promising to compare these data sources to real measurements, as we originally intended.

Future work could evaluate the effects of emulated devices on the quality of the measured data and, if negligible, scale up measurements to get an even better picture of ad targeting. Especially on iOS, we could not observe statistically significant targeting effects, which might be a product of our limited sample size. Additionally, our experiments can be extended to compare more smartphone manufacturers and OS versions.

Another interesting aspect of future work is Google's linking of services. We limited our experiments to the app stores for comparability reasons. A thorough study of how Google combines data from all its services to provide advertisements might shed more light on the mobile advertisement ecosystem by the example of a very large advertisement platform.

10 Conclusions

In this work, we provide insights into the advertisement and recommender systems of Apple's App Store on iOS and Google's Play Store on Android. These closed-source platforms are tightly integrated into the respective mobile OSs and classified as *gatekeepers* by the EU, and thus, cannot be avoided by most consumers.

To the best of our knowledge, researchers have not yet studied these mobile app stores' advertisement personalization and transparency. To address this research gap, we conduct large-scale measurements to perform a first analysis of those systems. In this process, we identify deficiencies in their transparency implementations of current EU regulations, present significant differences between Google's advertisement and recommender system, and share our experiences examining closed-source infrastructures to identify future improvements to app stores and regulations.

We provide our open-source methodology and dataset to enable researchers to examine those and similar systems to ensure user privacy in the future. From a legislative perspective, we support the idea of actionable transparency reports for personalization-based advertisement and recommender systems.

Disclosure Process

We shared our results regarding advertisement transparency defects with Apple and Google in February 2025. We received a detailed reply from Apple in March 2025 and incorporated their statements in Section 7. Google only provided an automated response and did not comment on our inquiry. Further, we reported our findings to the European Commission's DSA enforcement team.

Ethical Considerations

Apple Ads and Google App Campaigns only charge advertisers on CPI and CPT models (see Section 3). We have designed our experiments to never click on an ad within the app stores, and hence, have not caused any monetary loss for advertisers.

While we did exceed ordinary usage of the respective app stores, our experiments avoid unnecessary server queries. For example, we download app information only once per ad, even if the same ad emerges multiple times. Additionally, we assume that even the intense usage of two App Store and two Play Store instances is negligible, in contrast to Apple's and Google's worldwide traffic.

Our experiments did not involve any human beings and, therefore, did not need the approval of our institutional review board.

Availability

We published our source code on https://github.com/seemoo-lab/appstore-ad-tools and our dataset on https://doi.org/10.5281/zenodo.17037785.

Acknowledgments

We especially thank Tahireh Panahi from the University of Kassel for her insightful advice on the legal aspects of our work. Further, we thank the anonymous reviewers for their helpful suggestions. This work has been co-funded by the Federal Ministry of Education and Research of Germany in the project Open6GHub (grant number: 16KISK014) and the German Research Foundation (DFG) in the project CRUST (grant number: 503199853).

References

[1] Pushkal Agarwal, Sagar Joglekar, Panagiotis Papadopoulos, Nishanth Sastry, and Nicolas Kourtellis. 2020. Stop Tracking Me Bro! Differential Tracking of User Demographics on Hyper-Partisan Websites. In *Proceedings of The Web Conference 2020 (WWW '20)*. Association for Computing Machinery, New York, NY, USA, 1479–1490. https://doi.org/10.1145/3366423.3380221

[2] Muhammad Ali, Piotr Sapiezynski, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2021. Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM '21)*. Association for Computing Machinery, New York, NY, USA, 13–21. https://doi.org/10.1145/3437963.3441801

[3] Android. 2025. Android Help - Learn about the Android Device Configuration Service. Retrieved February 28, 2025 from https://support.google.com/android/answer/9021432

[4] Apple. 2023. Apple Advertising Policies. Retrieved February 28, 2025 from https://searchads.apple.com/policies

[5] Apple. 2024. Ad Repository API. Retrieved February 28, 2025 from https://developer.apple.com/support/downloads/Ad-Repository.pdf

[6] Apple. 2024. Apple Advertising & Privacy. Retrieved February 28, 2025 from https://www.apple.com/legal/privacy/data/en/apple-advertising/

[7] Apple. 2024. Apple Privacy Policy. Retrieved February 28, 2025 from https://www.apple.com/legal/privacy/en-ww/

[8] Apple. 2024. If an App Asks to Track Your Activity. Retrieved February 28, 2025 from https://support.apple.com/en-us/102420

[9] Apple. 2025. Ad Placement Options. Retrieved February 28, 2025 from https://ads.apple.com/app-store/help/ad-placements/0081-ad-placement-options

[10] Apple. 2025. Ad Repository. Retrieved February 28, 2025 from https://adrepository.apple.com/

[11] Apple. 2025. Ad Respoitory API. Retrieved February 28, 2025 from https://adrepository.apple.com/api/v1/

[12] Apple. 2025. App Store & Privacy. Retrieved February 28, 2025 from https://www.apple.com/legal/privacy/data/en/app-store/

[13] Apple. 2025. Apple Configurator – Benutzerhandbuch für den Mac. Retrieved February 28, 2025 from https://support.apple.com/de-de/guide/apple-configurator-mac/welcome/mac

[14] Apple. 2025. Apple Privacy Labels. Retrieved February 28, 2025 from https://www.apple.com/privacy/labels/

[15] Apple. 2025. Apple Search Ads. Retrieved February 28, 2025 from https://searchads.apple.com

[16] Apple. 2025. Commissions, Fees, and Taxes - App Store Connect - Help - Apple Developer. Retrieved February 28, 2025 from https://developer.apple.com/help/app-store-connect/distributing-apps-in-the-european-union/commissions-fees-and-taxes

[17] Apple. 2025. Compare Apple Search Ads Solutions. Retrieved February 28, 2025 from https://ads.apple.com/app-store/help/apple-search-ads-basic/0001-compare-apple-search-ads-solutions

[18] Apple. 2025. Data & Privacy. Retrieved February 28, 2025 from https://www.apple.com/legal/privacy/data/

[19] Apple. 2025. DSA Transparency Report. Retrieved May 21, 2025 from https://www.apple.com/legal/dsa/transparency/eu/app-store/2502/

[20] Apple. 2025. Get to Know Developer Mode | Documentation. Retrieved February 28, 2025 from https://wwdcnotes.com/documentation/wwdcnotes/wwdc22-110344-get-to-know-developer-mode/

[21] Apple. 2025. Getting Started with Web Distribution in the EU - Support - Apple Developer. Retrieved February 28, 2025 from https://developer.apple.com/support/web-distribution-eu/

[22] Apple. 2025. How Apple Uses Your Personal Data. Retrieved February 28, 2025 from https://privacy.apple.com/data/privacyinfo

[23] Apple. 2025. Modify Audience Settings. Retrieved February 28, 2025 from https://ads.apple.com/app-store/help/ad-groups/0021-modify-audience-settings

[24] Apple. 2025. Xcode - Testing. Retrieved February 28, 2025 from https://developer.apple.com/documentation/xcode/testing

[25] Paul Barford, Igor Canadi, Darja Krushevskaja, Qiang Ma, and S. Muthukrishnan. 2014. Adscape: Harvesting and Analyzing Online Display Ads. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14)*. Association for Computing Machinery, New York, NY, USA, 597–608. https://doi.org/10.1145/2566486.2567992

[26] Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson, and Christo Wilson. 2016. Tracing Information Flows Between Ad Exchanges Using Retargeted Ads. In *25th USENIX Security Symposium (USENIX Security 16)*. 481–496.

[27] Reuben Binns, Ulrik Lyngs, Max Van Kleek, Jun Zhao, Timothy Libert, and Nigel Shadbolt. 2018. Third Party Tracking in the Mobile Ecosystem. In *Proceedings of the 10th ACM Conference on Web Science*. ACM, Amsterdam Netherlands, 23–31. https://doi.org/10.1145/3201064.3201089

[28] Theodore Book and Dan S. Wallach. 2015. An Empirical Study of Mobile Ad Targeting. arXiv:1502.06577 [cs]

[29] Justin Brookman, Phoebe Rouge, Aaron Alva, and Christina Yeung. 2017. Cross-Device Tracking: Measurement and Disclosures. *Proceedings on Privacy Enhancing Technologies* (2017).

[30] Paolo Calciati, Konstantin Kuznetsov, Alessandra Gorla, and Andreas Zeller. 2020. Automatically Granted Permissions in Android Apps: An Empirical Study on Their Prevalence and on the Potential Threats for Privacy. In *Proceedings of the 17th International Conference on Mining Software Repositories (MSR '20)*. Association for Computing Machinery, New York, NY, USA, 114–124. https://doi.org/10.1145/3379597.3387469

[31] Juan Miguel Carrascosa, Jakub Mikians, Ruben Cuevas, Vijay Erramilli, and Nikolaos Laoutaris. 2015. I Always Feel like Somebody's Watching Me: Measuring Online Behavioural Advertising. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT '15)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/2716281.2836098

[32] European Commission. 2023. DSA: Very Large Online Platforms and Search Engines. Retrieved February 28, 2025 from https://ec.europa.eu/commission/presscorner/detail/en/ip_23_2413

[33] European Commission. 2024. Gatekeepers. Retrieved February 28, 2025 from https://digital-markets-act.ec.europa.eu/gatekeepers_en

[34] Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2015. Automated Experiments on Ad Privacy Settings. *Proceedings on Privacy Enhancing Technologies* (2015).

[35] Alex Davidson, Ian Goldberg, Nick Sullivan, George Tankersley, and Filippo Valsorda. 2018. Privacy Pass: Bypassing Internet Challenges Anonymously. *Proceedings on Privacy Enhancing Technologies* 2018, 3 (June 2018), 164–180. https://doi.org/10.1515/popets-2018-0026

[36] European Commission. 2017. Antitrust: Commission Fines Google €2.42 Billion for Abusing Dominance as Search Engine by Giving Illegal Advantage to Own Comparison Shopping Service.

[37] European Commission. 2018. Antitrust: Commission Fines Google €4.34 Billion for Illegal Practices Regarding Android Mobile Devices to Strengthen Dominance of Google's Search Engine.

[38] European Commission. 2019. Antitrust: Commission Fines Google €1.49 Billion for Abusive Practices in Online Advertising.

[39] European Commission. 2024. Commission Fines Apple over €1.8 Billion over Abusive App Store Rules for Music Streaming Providers.

[40] European Commission. 2025. Commission Finds Apple and Meta in Breach of the Digital Markets Act.

[41] European Parliament, Council of the European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation).

[42] European Parliament, Council of the European Union. 2022. Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on Contestable and Fair Markets in the Digital Sector and Amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).

[43] European Parliament, Council of the European Union. 2022. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and Amending Directive 2000/31/EC (Digital Services Act).

[44] Bundeszentrale für politische Bildung. 2022. Ausländische Bevölkerung nach Staatsangehörigkeit. Retrieved February 28, 2025 from https://www.bpb.de/kurz-knapp/zahlen-und-fakten/soziale-situation-in-deutschland/61631/auslaendische-bevoelkerung-nach-staatsangehoerigkeit/

[45] Eleni Gkiouzepi, Athanasios Andreou, Oana Goga, and Patrick Loiseau. 2023. Collaborative Ad Transparency: Promises and Limitations. In *2023 IEEE Symposium on Security and Privacy (SP)*. 2639–2657. https://doi.org/10.1109/SP46215.2023.10179448

[46] Phillip I. Good. 2000. *Permutation tests : a practical guide to resampling methods for testing hypotheses* (2. ed. ed.). Springer, New York. XVI, 270 S. pages.

[47] Google. 2022. Authorized Buyers - Protos & Reference Data. Retrieved May 21, 2025 from https://developers.google.com/authorized-buyers/rtb/data

[48] Google. 2022. Protos & Reference Data | Real-time Bidding. Retrieved February 28, 2025 from https://developers.google.com/authorized-buyers/rtb/data

[49] Google. 2024. Google AdMob - Earn more revenue with your apps. Retrieved February 28, 2025 from https://admob.google.com/home/

[50] Google. 2024. Google Cloud - google_ads_transparency_center. Retrieved February 28, 2025 from https://adstransparency.google.com/?region=DE

[51] Google. 2024. Google Privacy & Terms - Privacy Policy. Retrieved February 28, 2025 from https://policies.google.com/privacy

[52] Google. 2025. About DMA & Your Linked Services - Google Account Help. Retrieved February 28, 2025 from https://support.google.com/accounts/answer/14202510?hl=en

[53] Google. 2025. Ads Transparency Center. Retrieved February 28, 2025 from https://adstransparency.google.com/?region=DE

[54] Google. 2025. Age Requirements on Google Accounts - Google Account Help. Retrieved February 28, 2025 from https://support.google.com/accounts/answer/1350409

[55] Google. 2025. Android Help - Manage Your Ad Privacy Settings on Android. Retrieved February 28, 2025 from https://support.google.com/android/answer/13720755?hl=en#zippy=%2Cmanage-ad-topics%2Cmanage-app-suggested-ads

[56] Google. 2025. Android Open Accessory 2.0 Protocol. Retrieved February 28, 2025 from https://source.android.com/docs/core/interaction/accessories/aoa2

[57] Google. 2025. Developer Policy Center. Retrieved February 28, 2025 from https://play.google/developer-content-policy/

[58] Google. 2025. Google Ads Transparency Center - BigQuery Public Data. Retrieved February 28, 2025 from https://console.cloud.google.com/marketplace/details/bigquery-public-data/google-ads-transparency-center?inv=1&invt=AbqnSw

[59] Google. 2025. How the Google Ads Auction Works. Retrieved February 28, 2025 from https://support.google.com/google-ads/answer/6366577?hl=en

[60] Google. 2025. My Ad Center Help - Limit Ads about Sensitive Topics on Google. Retrieved February 28, 2025 from https://support.google.com/My-Ad-Center-Help/answer/12155260?hl=en

[61] Google. 2025. Promote Your Mobile App with App Campaigns. Retrieved February 28, 2025 from https://ads.google.com/intl/en_us/home/campaigns/app-ads/

[62] Google. 2025. Protecting User Privacy. Retrieved February 28, 2025 from https://publicpolicy.google/

[63] Google. 2025. Responsible Data Practices.

[64] Google. 2025. Service Fees - Play Console Help. Retrieved February 28, 2025 from https://support.google.com/googleplay/android-developer/answer/112622?hl=en

[65] Google. 2025. Welcome to My Activity. Retrieved February 28, 2025 from https://myactivity.google.com/

[66] Google. 2025. Write Automated Tests with UI Automator | Test Your App on Android. Retrieved February 28, 2025 from https://developer.android.com/training/testing/other-components/ui-automator

[67] Apple Inc. 2025. App Review Guidelines. Retrieved February 28, 2025 from https://developer.apple.com/app-store/review/guidelines/

[68] Costas Iordanou, Nicolas Kourtellis, Juan Miguel Carrascosa, Claudio Soriente, Ruben Cuevas, and Nikolaos Laoutaris. 2019. Beyond Content Analysis: Detecting Targeted Ads via Distributed Counting. In *Proceedings of the 15th International Conference on Emerging Networking Experiments And Technologies (CoNEXT '19)*. Association for Computing Machinery, New York, NY, USA, 110–122. https://doi.org/10.1145/3359989.3365428

[69] Umar Iqbal, Pouneh Nikkhah Bahrami, Rahmadi Trimananda, Hao Cui, Alexander Gamero-Garrido, Daniel J. Dubois, David Choffnes, Athina Markopoulou, Franziska Roesner, and Zubair Shafiq. 2023. Tracking, Profiling, and Ad Targeting in the Alexa Echo Smart Speaker Ecosystem. In *Proceedings of the 2023 ACM on Internet Measurement Conference (IMC '23)*. Association for Computing Machinery, New York, NY, USA, 569–583. https://doi.org/10.1145/3618257.3624803

[70] Yiming Jing, Ziming Zhao, Gail-Joon Ahn, and Hongxin Hu. 2014. Morpheus: Automatically Generating Heuristics to Detect Android Emulators. In *Proceedings of the 30th Annual Computer Security Applications Conference (ACSAC '14)*. Association for Computing Machinery, New York, NY, USA, 216–225. https://doi.org/10.1145/2664243.2664250

[71] Arjaldo Karaj, Sam Macbeth, Rémi Berson, and Josep M. Pujol. 2019. WhoTracks.Me: Shedding Light on the Opaque World of Online Tracking. https://doi.org/10.48550/arXiv.1804.08959 arXiv:1804.08959 [cs]

[72] Konrad Kollnig, Anastasia Shuba, Max Van Kleek, Reuben Binns, and Nigel Shadbolt. 2022. Goodbye Tracking? Impact of iOS App Tracking Transparency and Privacy Labels. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 508–520. https://doi.org/10.1145/3531146.3533116

[73] Mathias Lecuyer, Guillaume Ducoffe, Francis Lan, Andrei Papancea, Theofilos Petsios, Riley Spahn, Augustin Chaintreau, and Roxana Geambasu. 2014. XRay: Enhancing the Web's Transparency with Differential Correlation. (2014).

[74] Mathias Lecuyer, Riley Spahn, Yannis Spiliopolous, Augustin Chaintreau, Roxana Geambasu, and Daniel Hsu. 2015. Sunlight: Fine-grained Targeting Detection at Scale with Statistical Confidence. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, Denver Colorado USA, 554–566. https://doi.org/10.1145/2810103.2813614

[75] Douglas J. Leith. 2021. Mobile Handset Privacy: Measuring the Data iOS and Android Send to Apple and Google. In *Security and Privacy in Communication Networks*, Joaquin Garcia-Alfaro, Shujun Li, Radha Poovendran, Hervé Debar, and Moti Yung (Eds.). Vol. 399. Springer International Publishing, Cham, 231–251. https://doi.org/10.1007/978-3-030-90022-9_12

[76] Or Levi, Sardar Hamidian, and Pedram Hosseini. 2020. Automatically Identifying Political Ads on Facebook: Towards Understanding of Manipulation via User Targeting. In *Disinformation in Open Online Media*, Max van Duijn, Mike Preuss, Viktoria Spaiser, Frank Takes, and Suzan Verberne (Eds.). Springer International Publishing, Cham, 95–106. https://doi.org/10.1007/978-3-030-61841-4_7

[77] Thomas Linden, Rishabh Khandelwal, Hamza Harkous, and Kassem Fawaz. 2020. The Privacy Policy Landscape After the GDPR. *Proceedings on Privacy Enhancing Technologies* 2020, 1 (Jan. 2020), 47–64. https://doi.org/10.2478/popets-2020-0004

[78] Haoyu Liu, Douglas J. Leith, and Paul Patras. 2023. Android OS Privacy Under the Loupe – A Tale from the East. In *Proceedings of the 16th ACM Conference on Security and Privacy in Wireless and Mobile Networks*. ACM, Guildford United Kingdom, 31–42. https://doi.org/10.1145/3558482.3581775

[79] Haoyu Liu, Paul Patras, and Douglas J. Leith. 2023. On the Data Privacy Practices of Android OEMs. *PLOS ONE* 18, 1 (Jan. 2023), e0279942. https://doi.org/10.1371/journal.pone.0279942

[80] Yannic Meier, Anne Oeldorf-Hirsch, and Nicole C. Krämer. 2024. Who Is Targeting Me? Privacy Perceptions of and Responses to Commercial and Political Targeted Advertising on Social Media. *Journal of Advertising* 53, 4 (Aug. 2024), 473–490. https://doi.org/10.1080/00913367.2023.2275776

[81] Ministry of Public Security of the People's Republic of China. 2019. National Name Report. Retrieved February 28, 2025 from https://www.mps.gov.cn/n2254098/n4904352/c6874655/content.html

[82] Reham Mohamed, Arjun Arunasalam, Habiba Farrukh, Jason Tong, Antonio Bianchi, and Z Berkay Celik. 2024. ATTention Please! An Investigation of the App Tracking Transparency Permission. (2024).

[83] Maaz Bin Musa and Rishab Nithyanand. 2022. ATOM: Ad-network Tomography. *Proceedings on Privacy Enhancing Technologies* (2022).

[84] Mysk. 2022. The App Store on Your iPhone Is Watching Your Every Move. Retrieved February 28, 2025 from https://www.youtube.com/watch?v=8JxvH80Rrcw

[85] Javier Parra-Arnau, Jagdish Prasad Achara, and Claude Castelluccia. 2017. MyAdChoices: Bringing Transparency and Control to Online Advertising. *ACM Transactions on the Web* 11, 1 (March 2017), 7:1–7:47. https://doi.org/10.1145/2996466

[86] Joel Reardon, Álvaro Feal, Primal Wijesekera, Amit Elazari Bar On, Narseo Vallina-Rodriguez, and Serge Egelman. 2019. 50 Ways to Leak Your Data: An Exploration of Apps' Circumvention of the Android Permissions System. In *28th USENIX Security Symposium (USENIX Security 19)*. 603–620.

[87] Selenium. 2025. Selenium. Retrieved February 28, 2025 from https://www.selenium.dev/

[88] Anastasia Shuba and Athina Markopoulou. 2020. NoMoATS: Towards Automatic Detection of Mobile Tracking. *Proceedings on Privacy Enhancing Technologies* 2020, 2 (April 2020), 45–66. https://doi.org/10.2478/popets-2020-0017

[89] Konstantinos Solomos, Panagiotis Ilia, Sotiris Ioannidis, and Nicolas Kourtellis. 2019. {TALON}: An Automated Framework for {Cross-Device} Tracking Detection. In *22nd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2019)*. 227–241.

[90] Statista. 2024. Biggest App Stores in the World 2024. Retrieved February 28, 2025 from https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/

[91] Statista. 2025. Mobile OS Market Share Worldwide 2009-2024. Retrieved February 28, 2025 from https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/

[92] Statistisches Bundesamt. 2000. Statistisches Jahrbuch.

[93] Joanna Strycharz and Bram Duivenvoorde. 2021. The Exploitation of Vulnerability through Personalised Marketing Communication: Are Consumers Protected? *Internet Policy Review* 10, 4 (Nov. 2021).

[94] Siming Su, Haoyu Wang, and Guoai Xu. 2021. Towards Understanding iOS App Store Search Advertising: An Explorative Study. In *2021 IEEE/ACM 8th International Conference on Mobile Software Engineering and Systems (MobileSoft)*. 40–51. https://doi.org/10.1109/MobileSoft52590.2021.00011

[95] Michael Carl Tschantz, Amit Datta, Anupam Datta, and Jeannette M. Wing. 2015. A Methodology for Information Flow Experiments. In *2015 IEEE 28th Computer Security Foundations Symposium*. 554–568. https://doi.org/10.1109/CSF.2015.40

[96] Turkish Statistical Institute. 2024. The Most Used Men Names by the Year of Birth and Order in the First Hundred Names. Retrieved February 28, 2025 from https://data.tuik.gov.tr/Bulten/DownloadIstatistikselTablo?p=/jWCDlUxCSaA9BFALPBb1hE4T74xvemaHSEv466skqdiXRZMkddauudaBwdsSmFm

[97] Imdad Ullah, Roksana Boreli, and Salil S. Kanhere. 2023. Privacy in Targeted Advertising on Mobile Devices: A Survey. *International Journal of Information Security* 22, 3 (June 2023), 647–678. https://doi.org/10.1007/s10207-022-00655-x

[98] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS '12)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/2335356.2335362

[99] Sebastian Zimmeck, Jie S. Li, Hyungtae Kim, Steven M. Bellovin, and Tony Jebara. 2017. A Privacy Analysis of Cross-device Tracking. In *26th USENIX Security Symposium (USENIX Security 17)*. 1391–1408.

[100] Shivani Zoting and Aditi Shivarkar. 2025. Mobile Application Market Size to Hit USD 1,103.48 Bn by 2034. Retrieved February 28, 2025 from https://www.precedenceresearch.com/mobile-application-market

A Appendix

Section A.1 provides information on our experiments and the personas we used. In Sections A.2 and A.3, we provide supplementary plots of our created dataset.

A.1 Persona and Experiment Details

Table 2 lists all conducted experiments. Table 3 specifies the parameters we use for our personas.

A.2 Jaccard Similarity Matrices

Figures 6 and 7 show the mean Jaccard similarity for our three types of personas. Following our methodology, data extracted on Android devices is split into ads and recommendations, while iOS data is split into the Today and Search tab.

A.3 Category Distribution Data

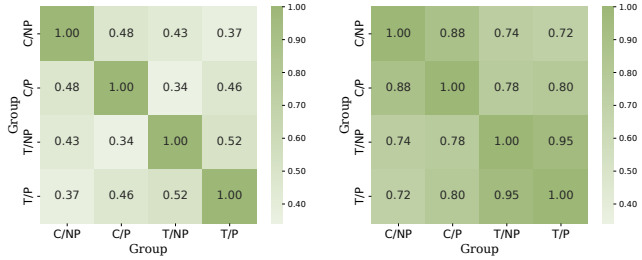
Figures 8 and 9 depict the complete category distribution data of all our main experiments. On Google's Play Store, the recommendation container was not always present when measuring ads, which resulted in fluctuations in the amount of recommended apps that we were able to obtain.

Table 2: Overview of all experiments and respective selected treatment and control personas.

Platform	Treatment Persona	Control Persona	Number of Repetitions	Experiment Description
Play Store (Android)	Gender Female	Gender Male	5	Testing the effect of names associated with a gender.
	Name – Chinese	Neutral	5	Testing the influence of a Chinese name.
	Name – Turkish	Neutral	5	Testing the influence of a Turkish name.
	Age Group 25	Age Group 50	3	Comparing a persona age of 25 to age 50.
	Age Group 25	Age Group 85	3	Comparing a persona age of 25 to age 85.
	Age Group 50	Age Group 85	3	Comparing a persona age of 50 to age 85.
	Shopping	Neutral	5	Persona installing different shopping related apps.
	Finance	Neutral	5	Persona installing different finance related apps.
	Parenting	Neutral	5	Persona installing different parenting related apps.
	Mental Health	Neutral	5	Persona installing different mental health related apps.
	Weight Loss	Neutral	5	Persona installing different weight-loss related apps.
	Alcohol sobriety	Neutral	5	Persona installing alcohol sobriety related apps.
	Gambling	Neutral	5	Persona installing different gambling related apps.
	Neutral	Neutral	5	Neutral persona with default parameters and no apps.
App Store (iOS)	Gender Female	Gender Male	5	Testing the effect of names associated with a gender.
	Name – Chinese	Neutral	5	Testing the influence of a Chinese name.
	Name – Turkish	Neutral	5	Testing the influence of a Turkish name.
	Age Group 25	Age Group 50	3	Comparing an account birthday age of 25 to age 50.
	Age Group 25	Age Group 85	3	Comparing an account birthday age of 25 to age 85.
	Age Group 50	Age Group 85	3	Comparing an account birthday age of 50 to age 85.
	Shopping	Neutral	5	Persona installing different shopping related apps.
	Finance	Neutral	5	Persona installing different finance related apps.
	Parenting	Neutral	5	Persona installing different parenting related apps.
	Mental Health	Neutral	5	Persona installing different mental health related apps.
	Weight Loss	Neutral	5	Persona installing different weight-loss related apps.
	Alcohol sobriety	Neutral	5	Persona installing alcohol sobriety related apps.
	Gambling	Neutral	5	Persona installing different gambling related apps.
	Neutral	Neutral	5	Neutral persona with default parameters and no apps.

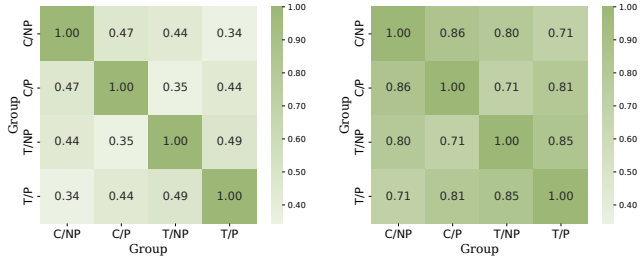
Table 3: Overview of our personas per platform.

Category	Persona	Name	Age	Associated Apps	#Accounts
Account parameter	Gender Female	Top 5 female first names + Müller	34	-	5
	Gender Male	Top 5 male first names + Müller	34		5
	Name – Turkish	Mehmet Yilmaz	34		5
	Name – Chinese	Zhang Wei	34		5
	Age Group 25	Alex Müller	25		3
	Age Group 50	Alex Müller	50		3
	Age Group 85	Alex Müller	85		3
Interest-based	Shopping	Alex Müller	34	Top 30 shopping apps	5
	Finance	Alex Müller	34	Top 30 finance apps	5
	Parenting	Alex Müller	34	30 curated parenting apps	5
	Mental Health	Alex Müller	34	30 curated mental well-being apps	5
	Weight Loss	Alex Müller	34	30 curated food tracker apps	5
	Alcohol sobriety	Alex Müller	34	Top 30 sobriety tracker apps	5
	Gambling	Alex Müller	34	Top 30 casino apps	5
Neutral	Neutral	Alex Müller	34	-	55



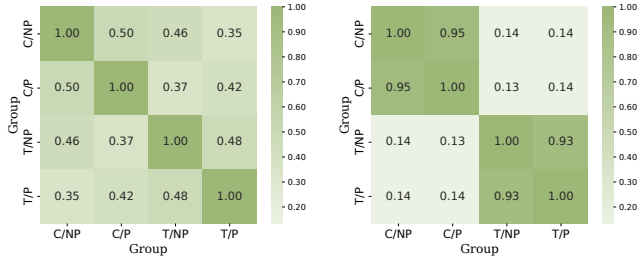
(a) Baseline ads

(b) Baseline recommendations



(c) Account param. ads

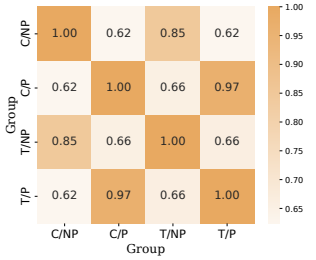
(d) Account param. recomms.



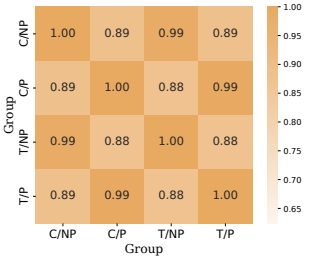
(e) Interest-based ads

(f) Interest-based recomms.

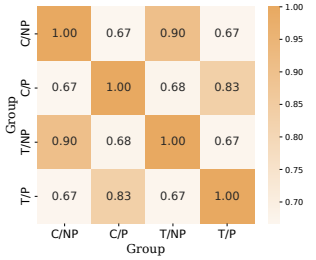
Figure 6: Mean Jaccard similarity for ad / recommendation name frequency on Android.



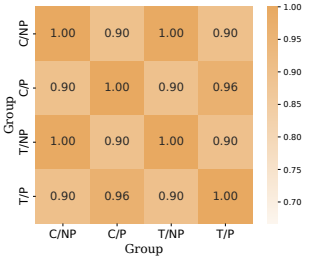
(a) Baseline ads (Search tab)



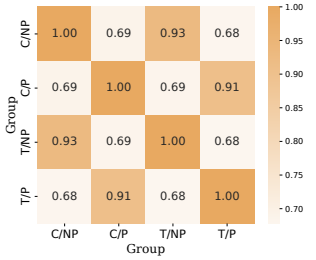
(b) Baseline ads (Today tab)



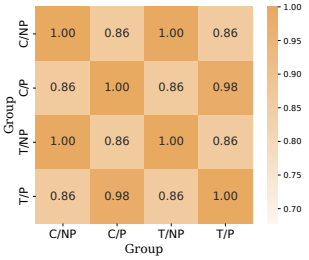
(c) Acc. param. ads (Search tab)



(d) Acc. param. ads (Today tab)

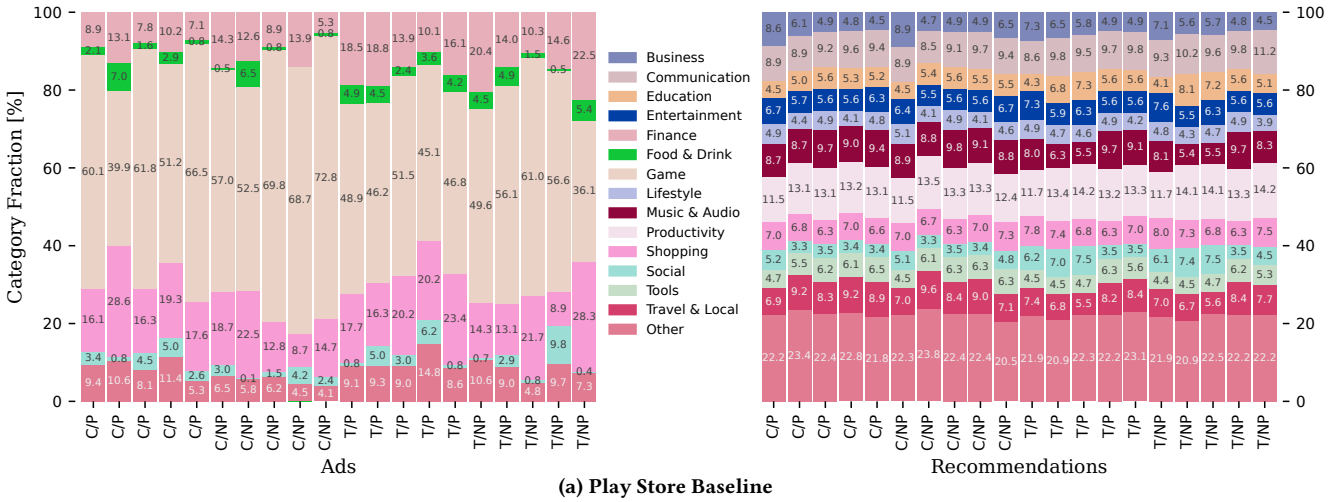


(e) Interest-based ads (Search)

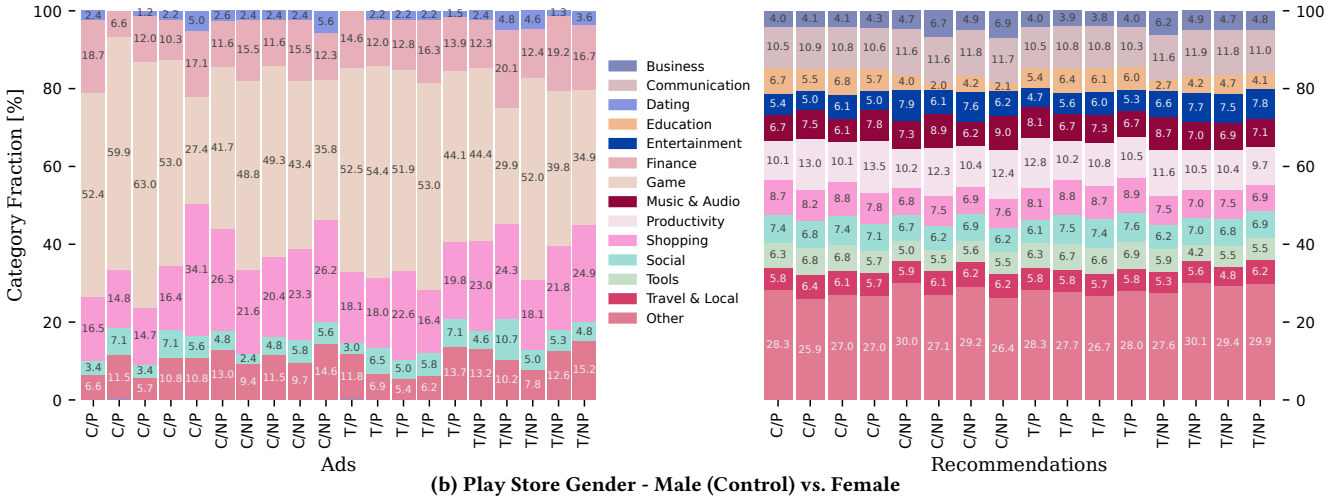


(f) Interest-based ads (Today)

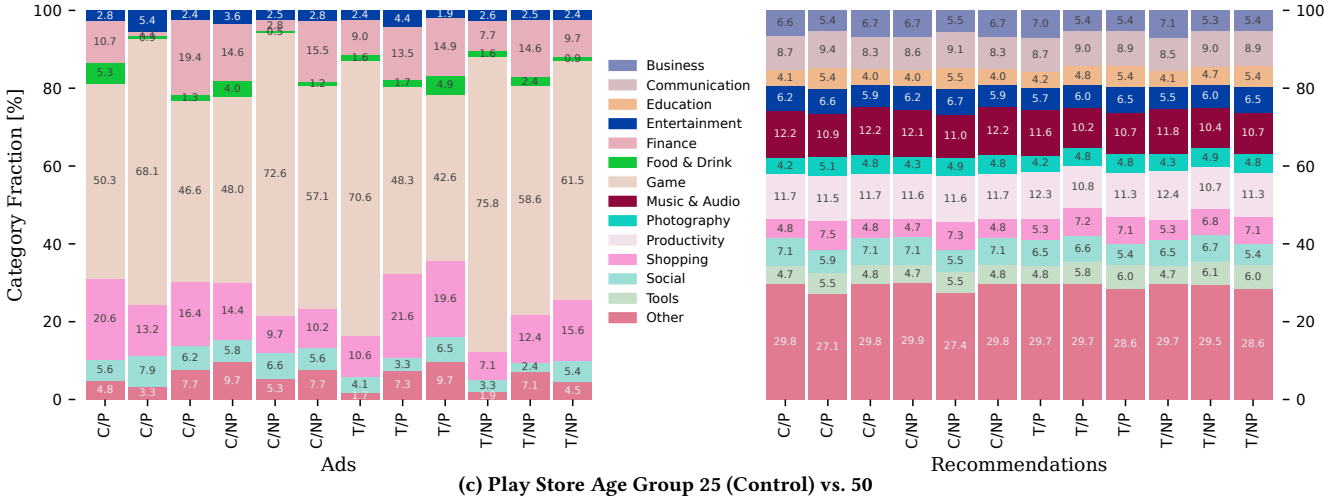
Figure 7: Mean Jaccard similarity for the ad name frequency on iOS.



(a) Play Store Baseline

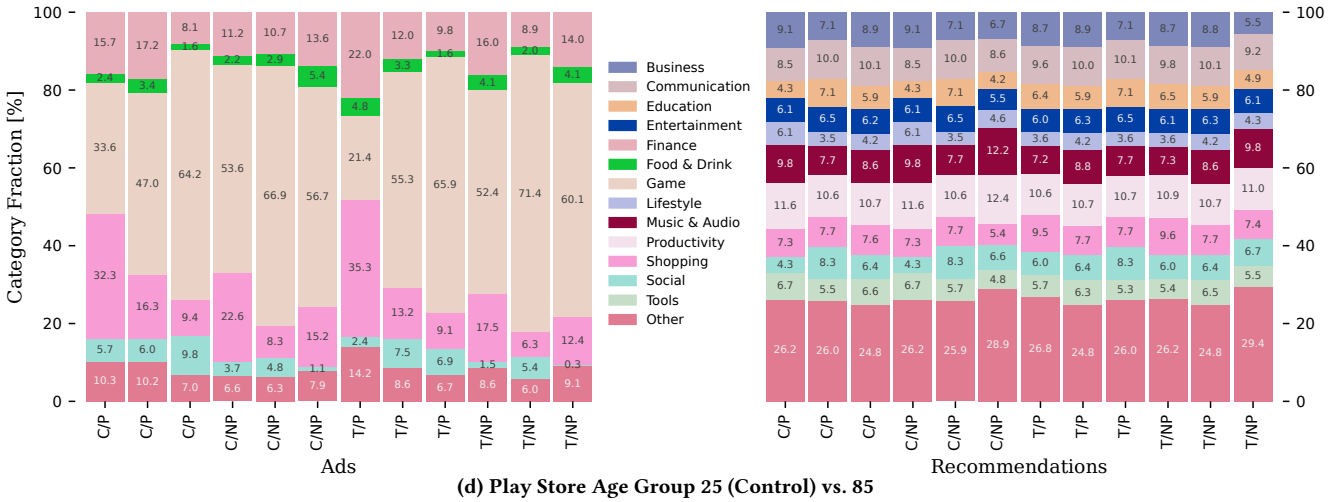


(b) Play Store Gender - Male (Control) vs. Female

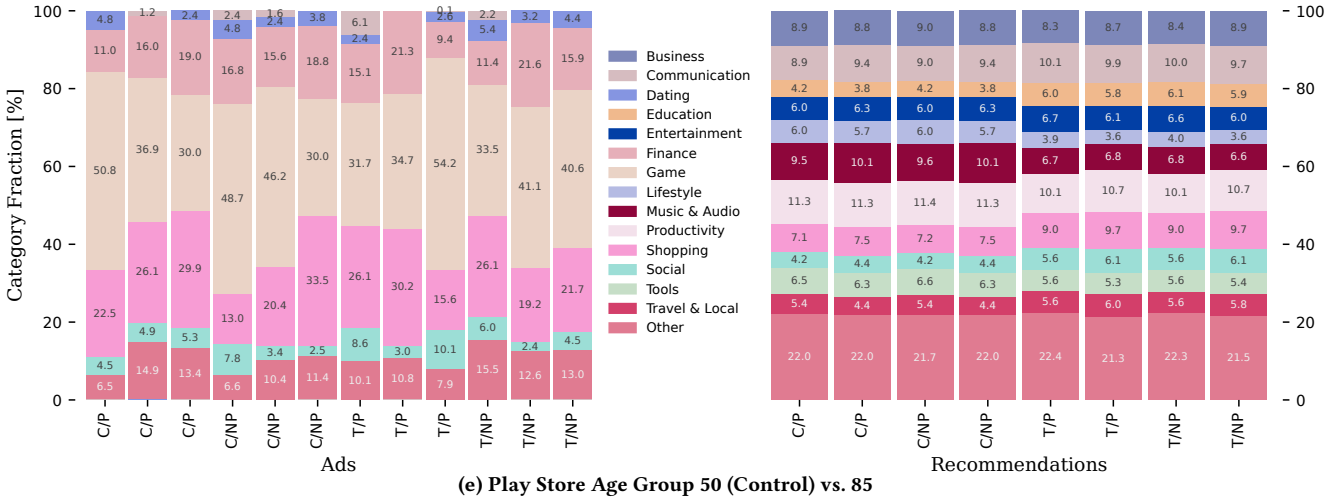


(c) Play Store Age Group 25 (Control) vs. 50

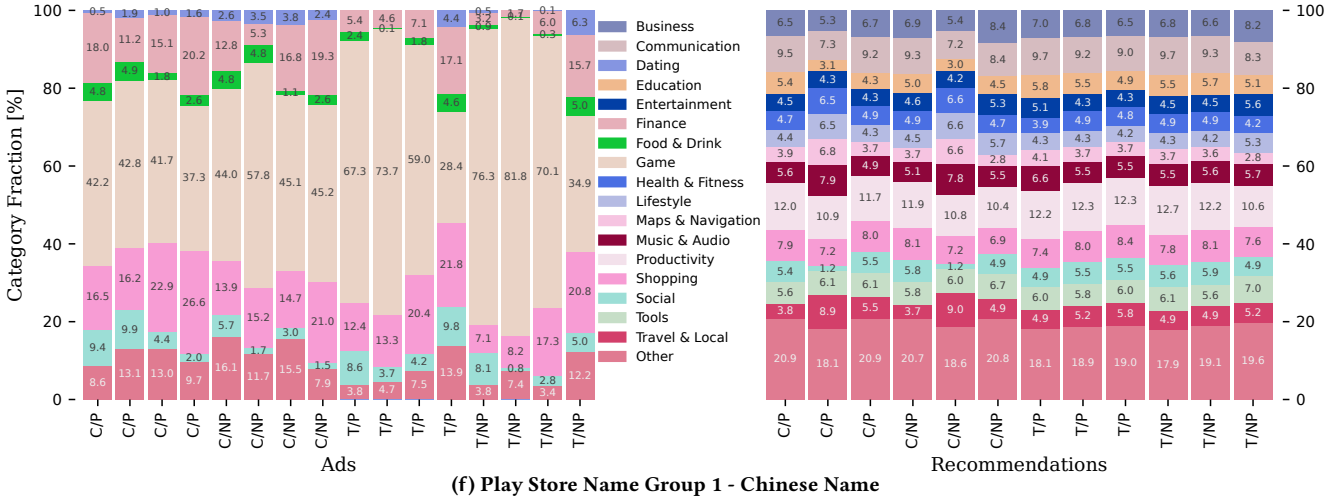
Figure 8: Category distribution of all Play Store measurements before and after deactivating personalized ads (C = Control, T = Treatment, P = Personalized, NP = Non-Personalized). "Other" captures all categories that are consistently below 5%.



(d) Play Store Age Group 25 (Control) vs. 85

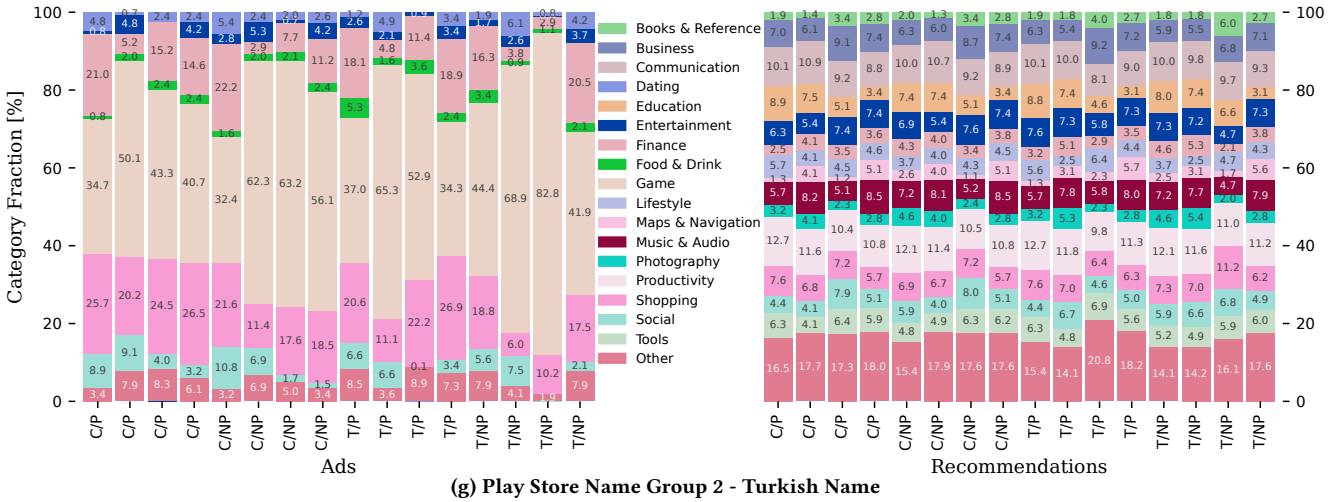


(e) Play Store Age Group 50 (Control) vs. 85

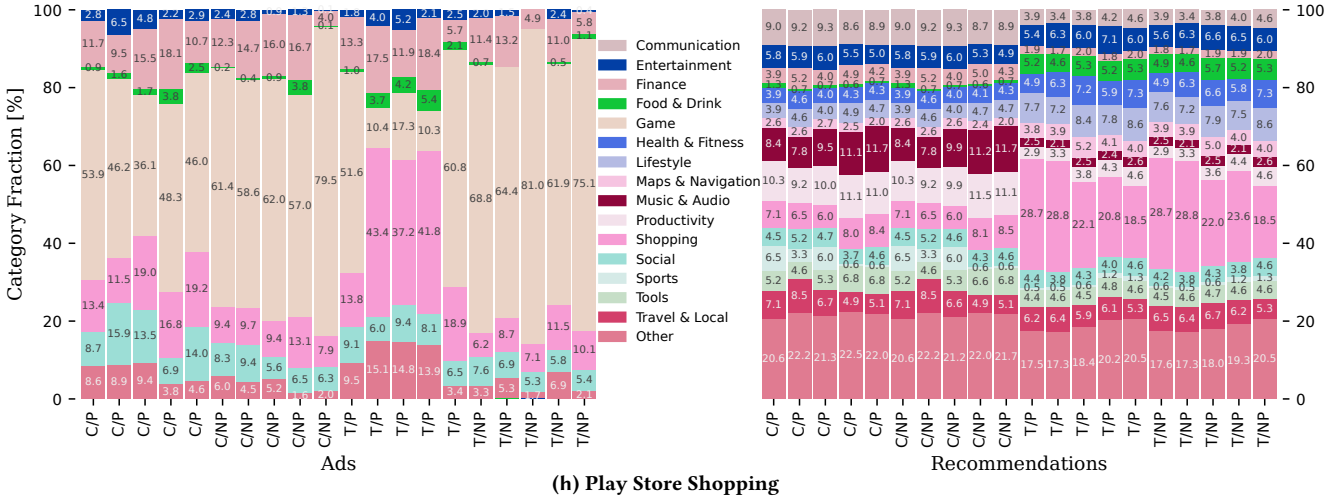


(f) Play Store Name Group 1 - Chinese Name

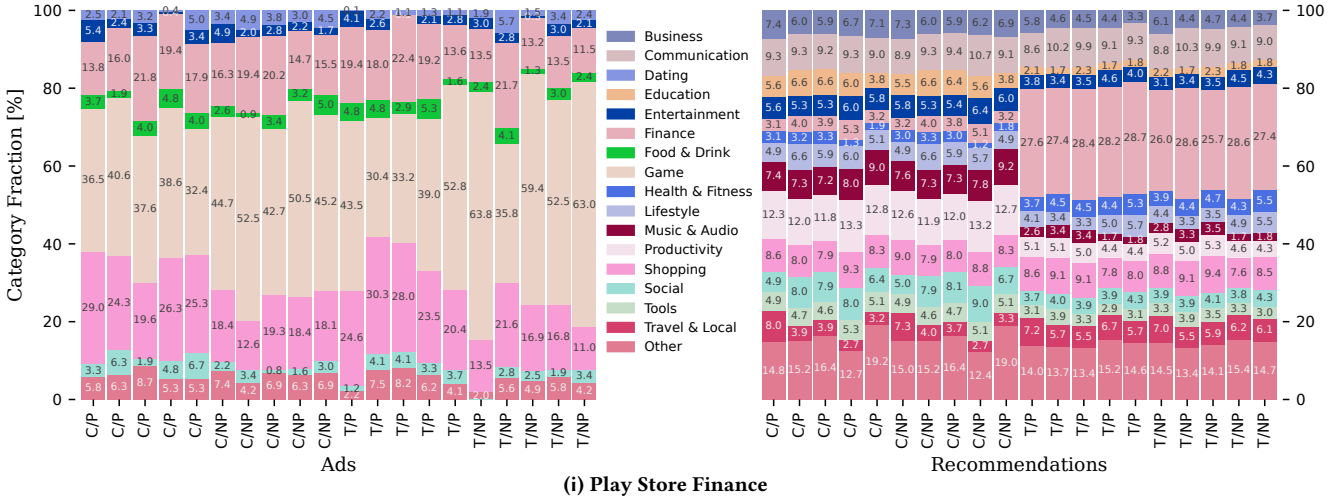
Figure 8: Category distribution of all Play Store measurements before and after deactivating personalized ads (C = Control, T = Treatment, P = Personalized, NP = Non-Personalized). "Other" captures all categories that are consistently below 5%.



(g) Play Store Name Group 2 - Turkish Name

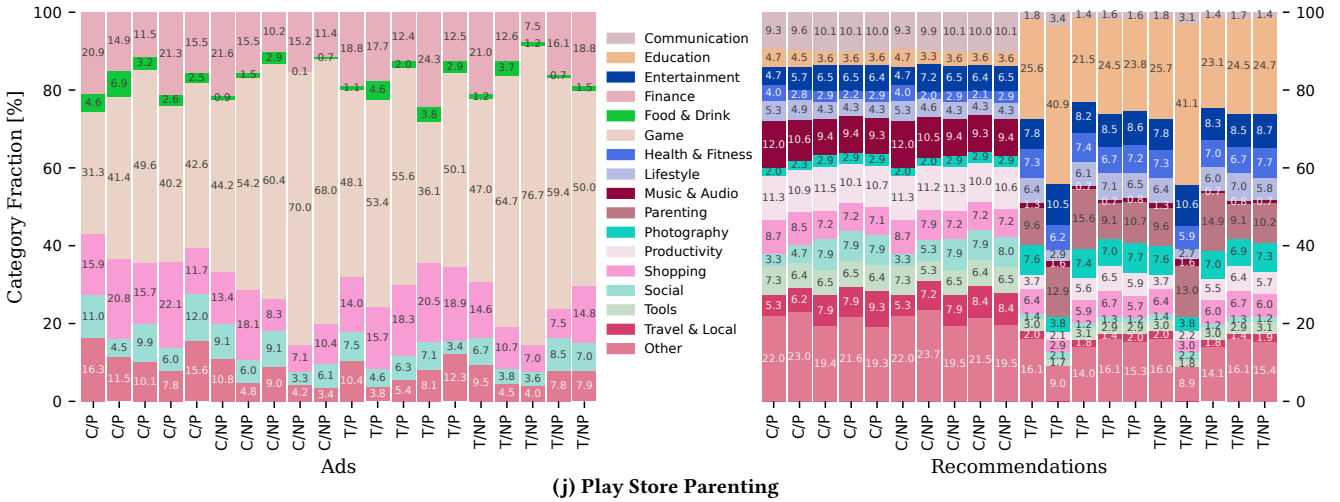


(h) Play Store Shopping

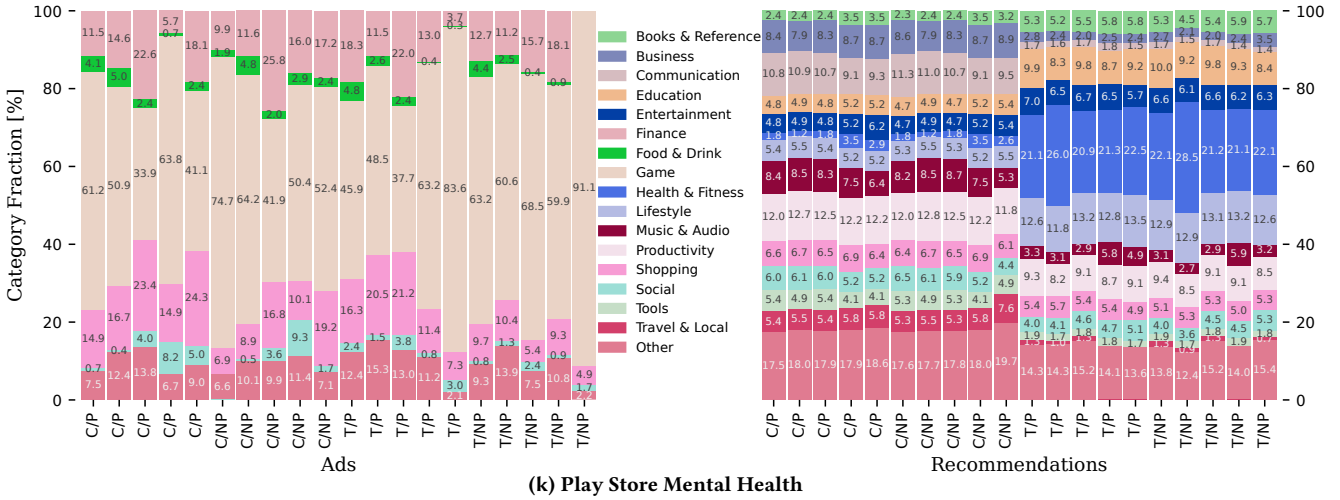


(i) Play Store Finance

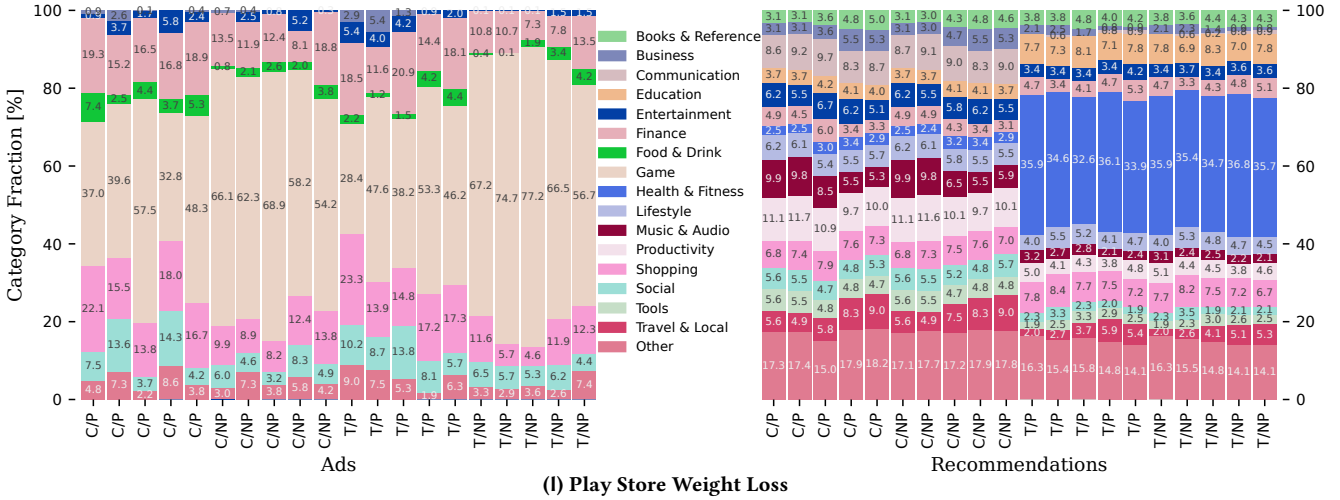
Figure 8: Category distribution of all Play Store measurements before and after deactivating personalized ads (C = Control, T = Treatment, P = Personalized, NP = Non-Personalized). "Other" captures all categories that are consistently below 5%.



(j) Play Store Parenting

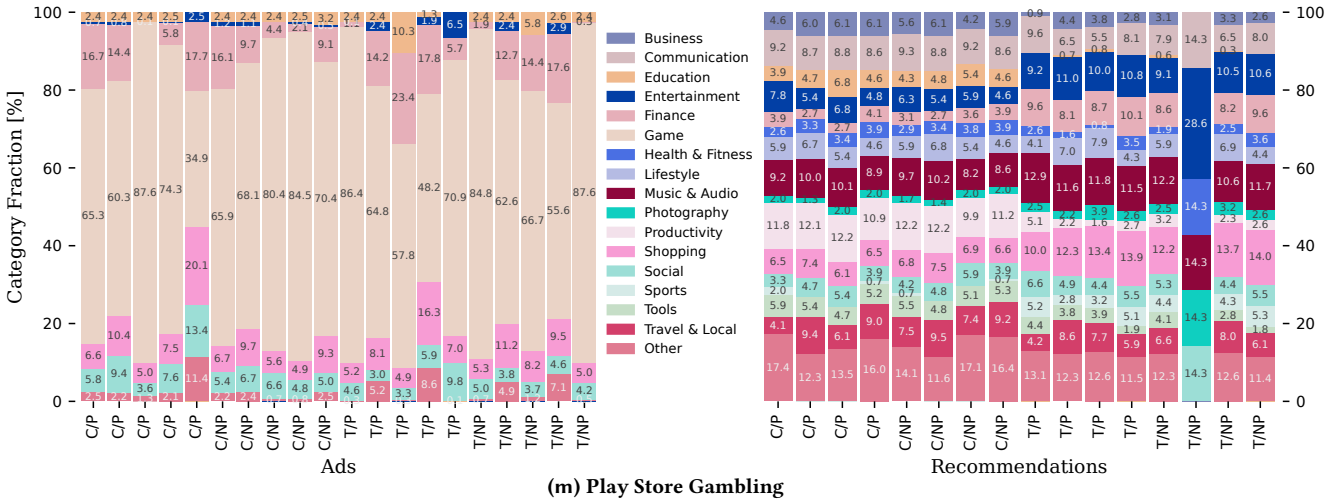


(k) Play Store Mental Health

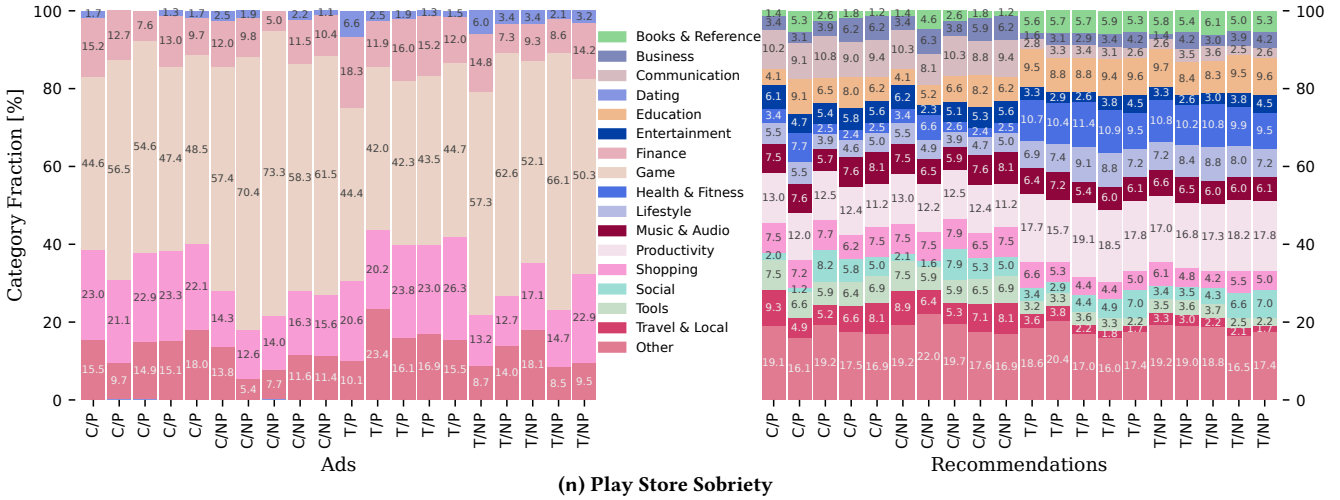


(l) Play Store Weight Loss

Figure 8: Category distribution of all Play Store measurements before and after deactivating personalized ads (C = Control, T = Treatment, P = Personalized, NP = Non-Personalized). "Other" captures all categories that are consistently below 5%.

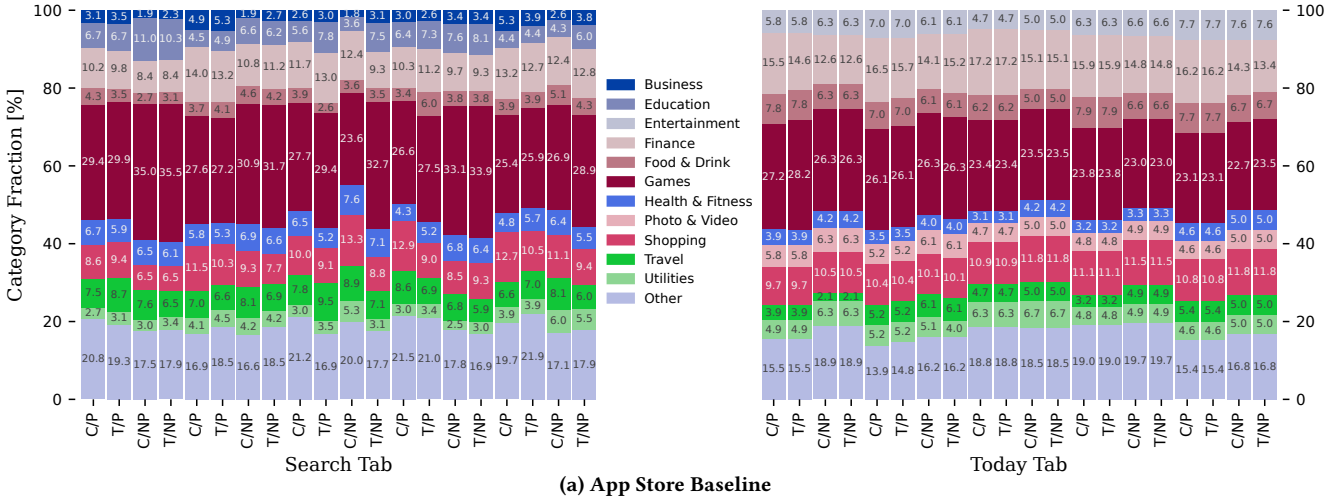


(m) Play Store Gambling

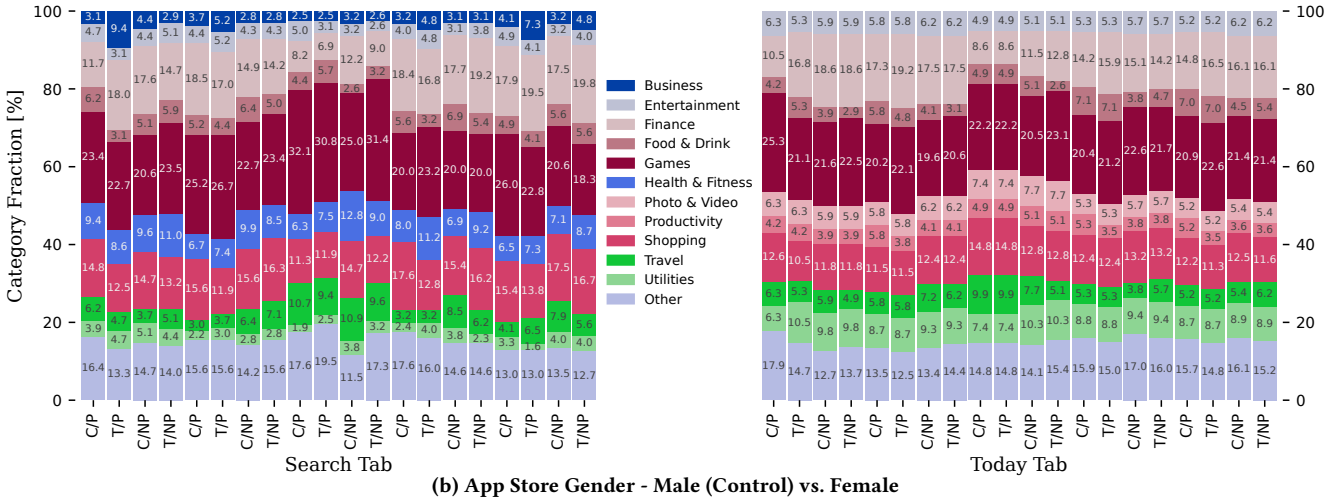


(n) Play Store Sobriety

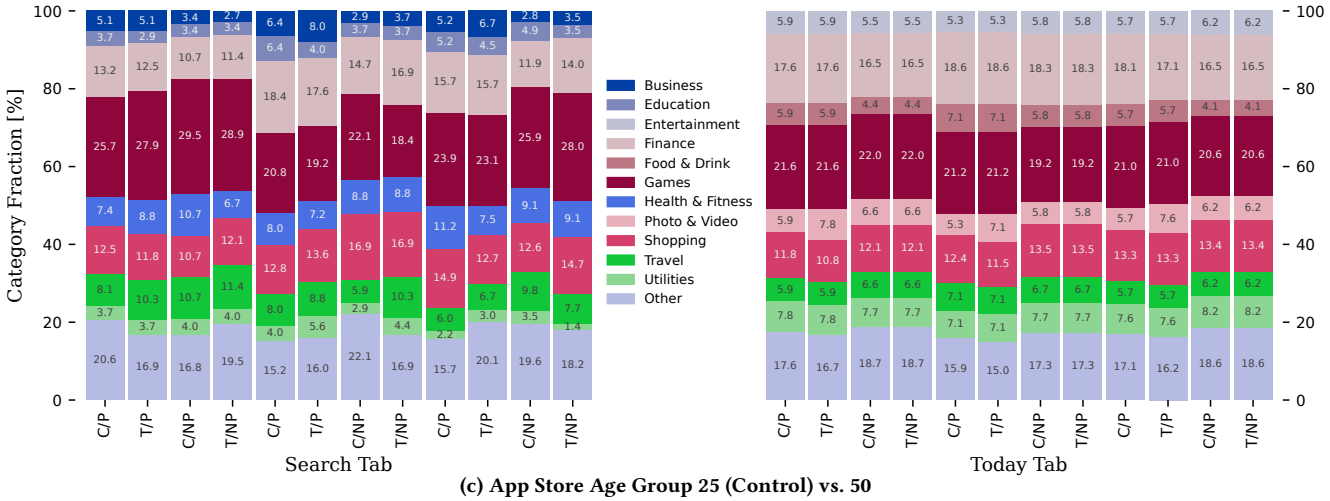
Figure 8: Category distribution of all Play Store measurements before and after deactivating personalized ads (C = Control, T = Treatment, P = Personalized, NP = Non-Personalized). "Other" captures all categories that are consistently below 5%.



(a) App Store Baseline

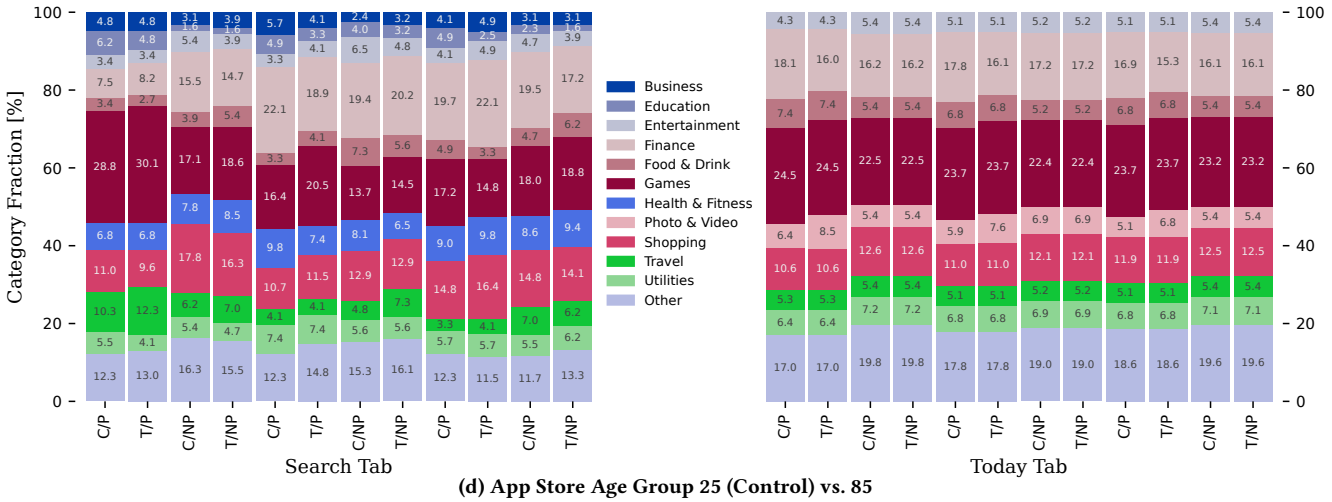


(b) App Store Gender - Male (Control) vs. Female

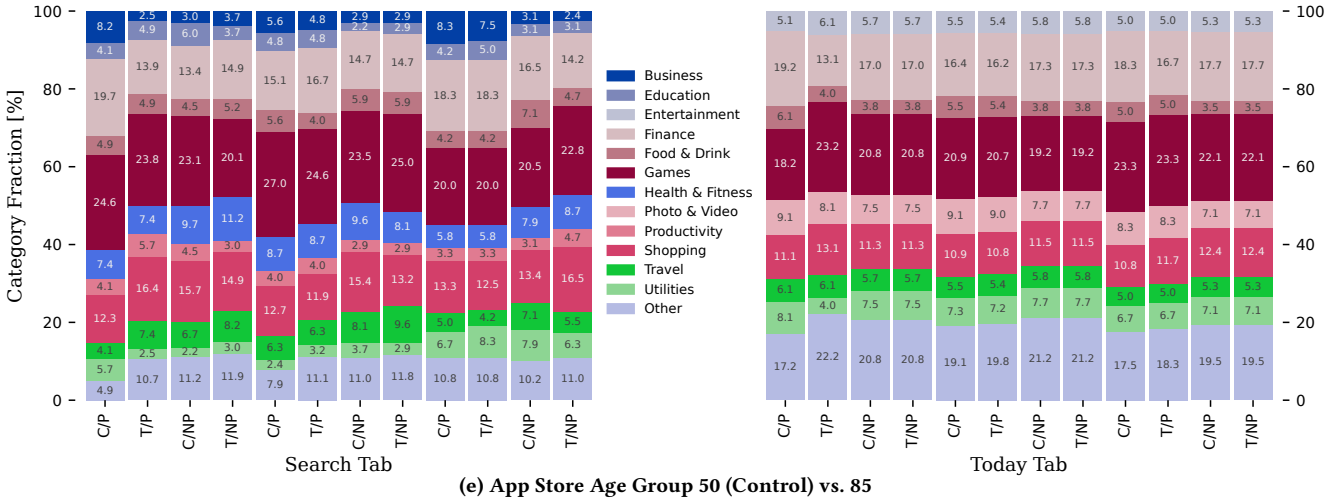


(c) App Store Age Group 25 (Control) vs. 50

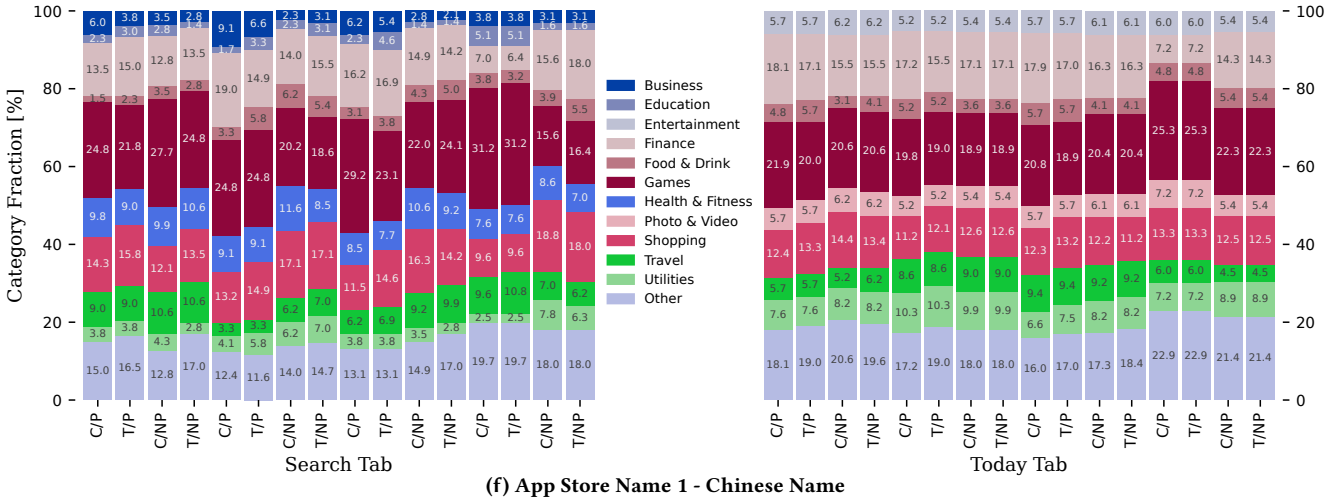
Figure 9: Category distribution of all App Store measurements before and after deactivating personalized ads (C = Control, T = Treatment, P = Personalized, NP = Non-Personalized). "Other" captures all categories that are consistently below 5%.



(d) App Store Age Group 25 (Control) vs. 85

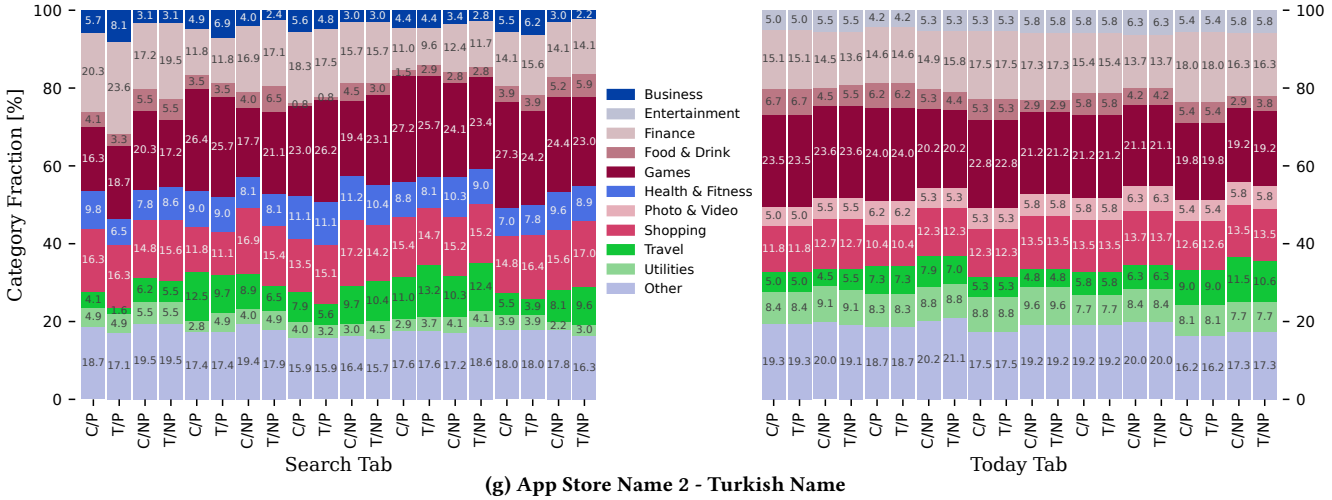


(e) App Store Age Group 50 (Control) vs. 85

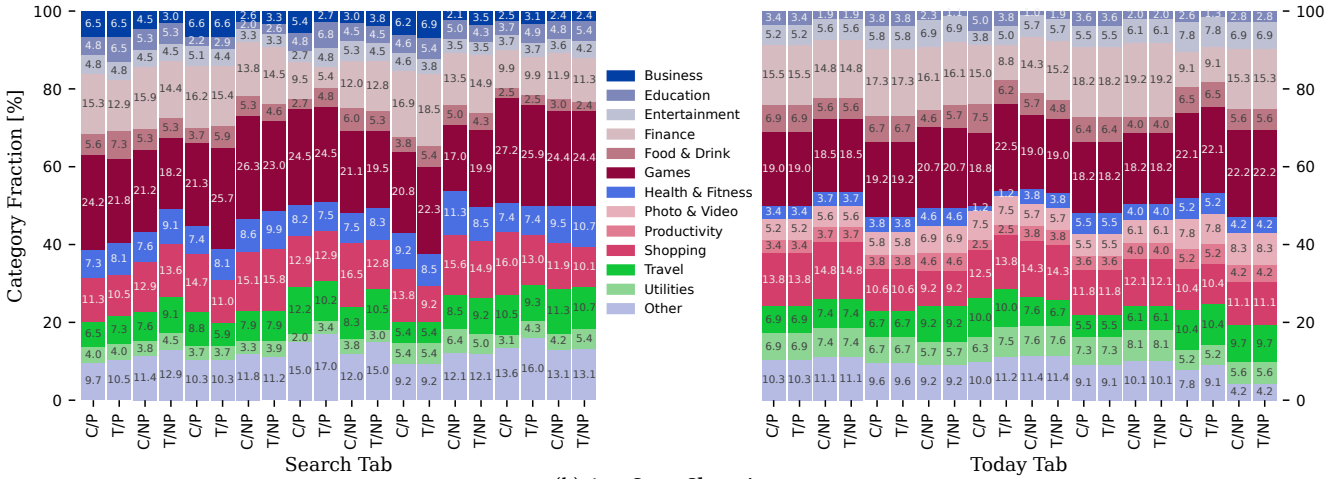


(f) App Store Name 1 - Chinese Name

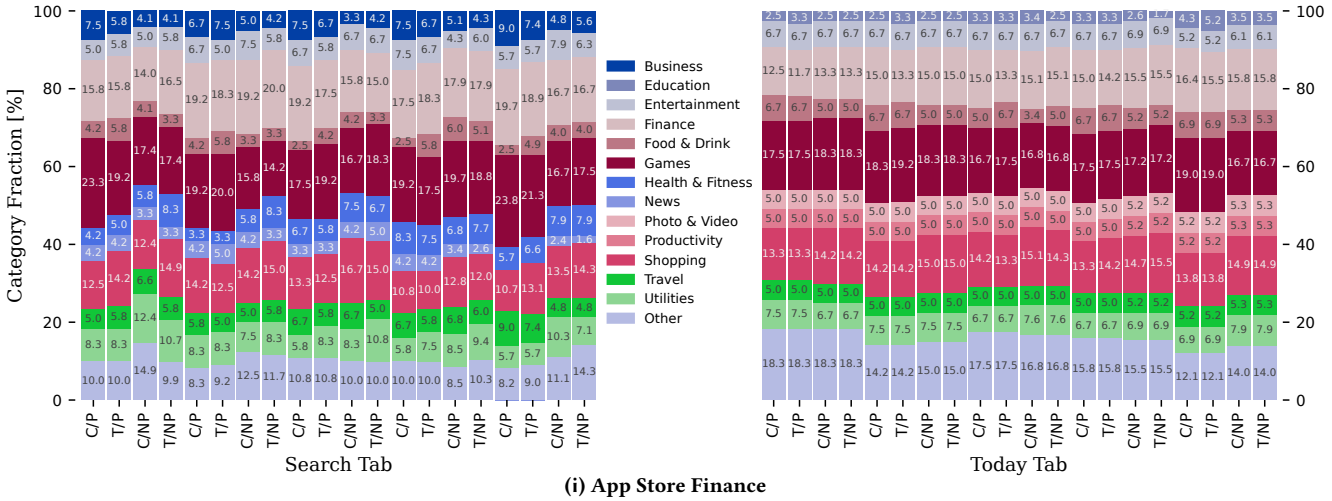
Figure 9: Category distribution of all App Store measurements before and after deactivating personalized ads (C = Control, T = Treatment, P = Personalized, NP = Non-Personalized). "Other" captures all categories that are consistently below 5%.



(g) App Store Name 2 - Turkish Name

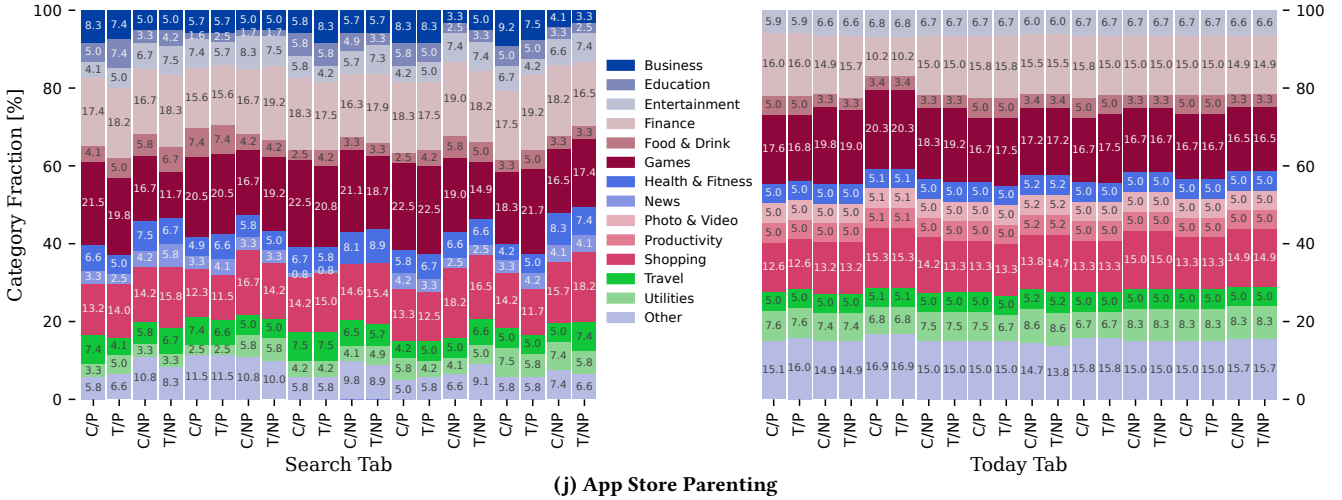


(h) App Store Shopping

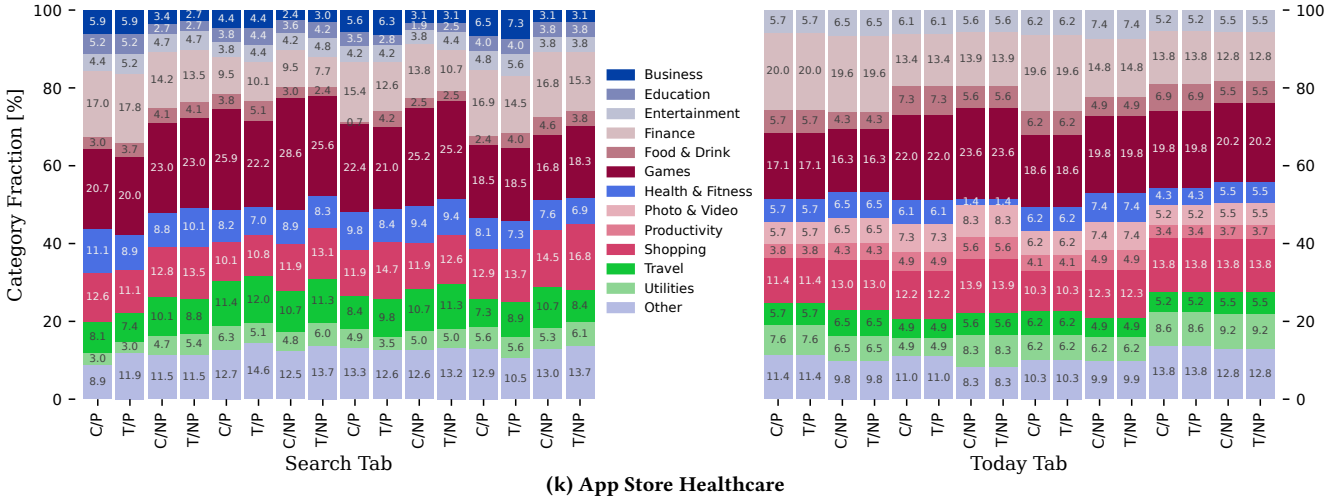


(i) App Store Finance

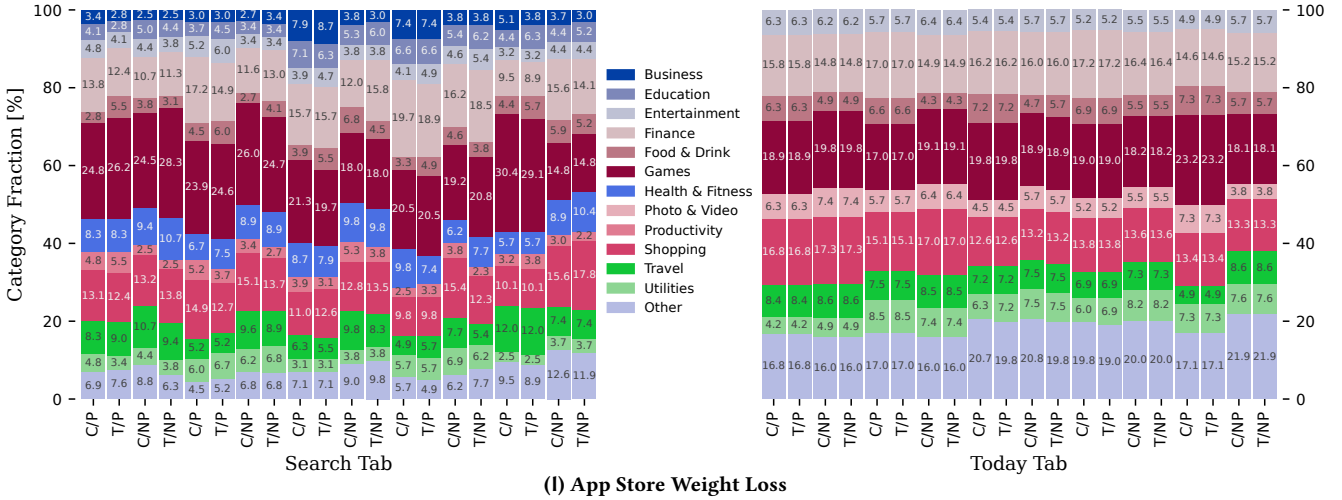
Figure 9: Category distribution of all App Store measurements before and after deactivating personalized ads (C = Control, T = Treatment, P = Personalized, NP = Non-Personalized). "Other" captures all categories that are consistently below 5%.



(j) App Store Parenting

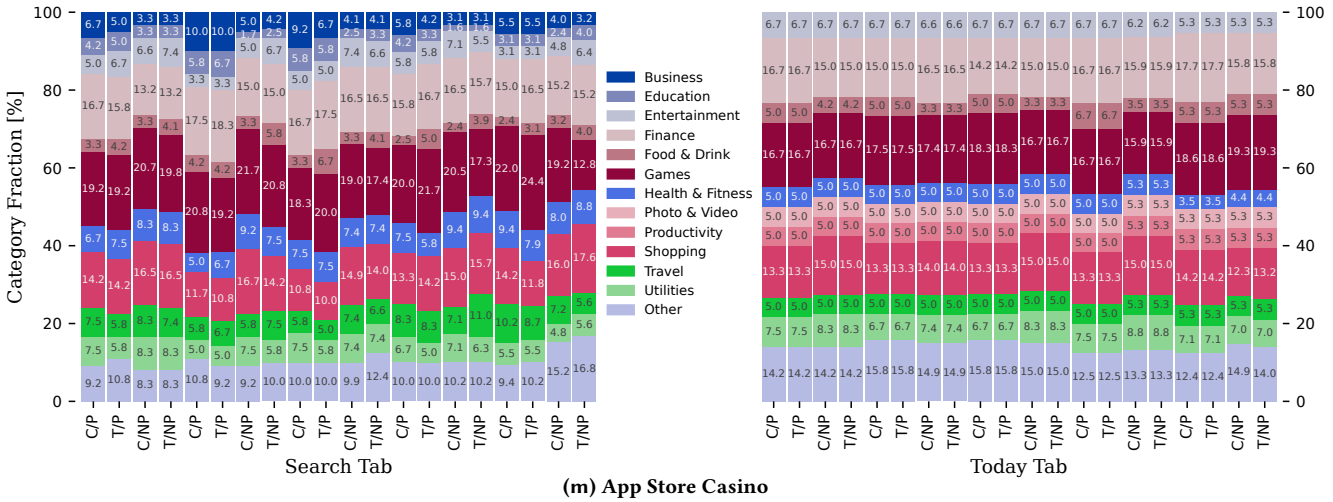


(k) App Store Healthcare

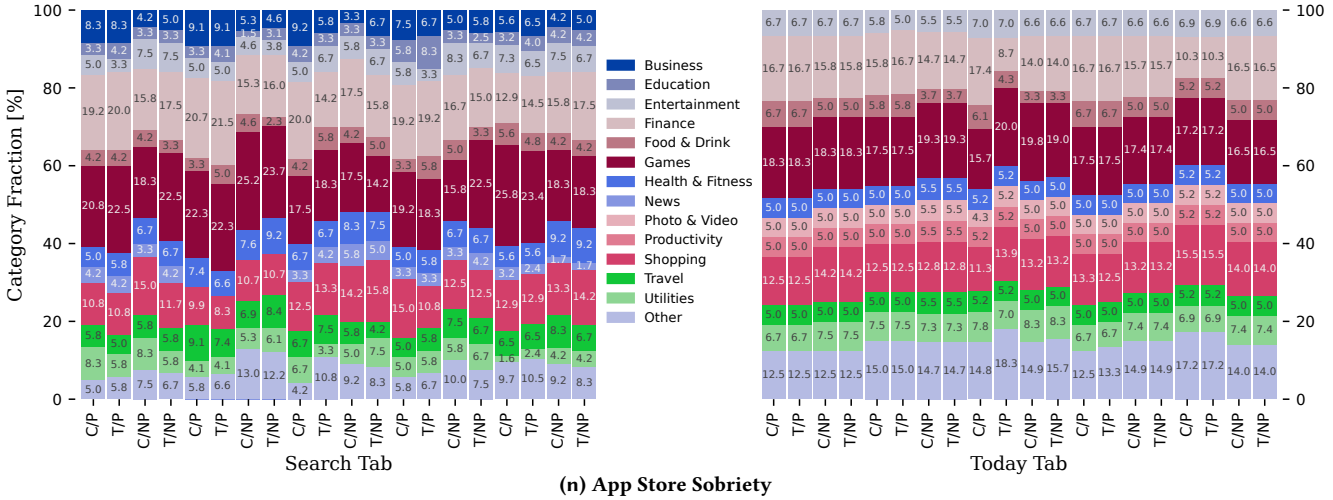


(l) App Store Weight Loss

Figure 9: Category distribution of all App Store measurements before and after deactivating personalized ads (C = Control, T = Treatment, P = Personalized, NP = Non-Personalized). "Other" captures all categories that are consistently below 5%.



(m) App Store Casino



(n) App Store Sobriety

Figure 9: Category distribution of all App Store measurements before and after deactivating personalized ads (C = Control, T = Treatment, P = Personalized, NP = Non-Personalized). "Other" captures all categories that are consistently below 5%.